Journal of Universal Computer Science, vol. 26, no. 4 (2020), 496-507
submitted: 11/10/2019, accepted: 5/2/2020, appeared: 28/4/2020 CC BY-ND 4.0

Real-Time Bot Detection from Twitter Using the Twitterbot+ Framework

Kheir Eddine Daouadi
(Faculty of Economics and Management, MIRACL, Sfax University, Sfax, Tunisia
khairi.informatique@gmail.com)

Rim Zghal Rebaï
(Higher Institute of Computer Science and Multimedia, MIRACL, Sfax University
Sfax, Tunisia
rim.zghal@gmail.com)

Ikram Amous
(National School of Electronic and Telecom, MIRACL, Sfax University, Sfax, Tunisia
ikram.amous@enetcom.usf.tn)

Abstract: Nowadays, bot detection from Twitter attracts the attention of several researchers around the world. Different bot detection approaches have been proposed as a result of these research efforts. Four of the main challenges faced in this context are the diversity of types of content propagated throughout Twitter, the problem inherent to the text, the lack of sufficient labeled datasets and the fact that the current bot detection approaches are not sufficient to detect bot activities accurately. We propose, Twitterbot+, a bot detection system that leveraged a minimal number of language-independent features extracted from one single tweet with temporal enrichment of a previously labeled datasets. We conducted experiments on three benchmark datasets with standard evaluation scenarios, and the achieved results demonstrate the efficiency of Twitterbot+ against the state-of-the-art. This yielded a promising accuracy results (>95%). Our proposition is suitable for accurate and real-time use in a Twitter data collection step as an initial filtering technique to improve the quality of research data.

Key Words: Twitter Bot Detection, Twitterbot+, Statistical-based Approach, Content-based Approach, Hybrid-based Approach, Temporal Enrichment
Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

1 Introduction

Nowadays, social media attract the attention of researchers, companies, politicians, and even governmental institutions. This gain of interest is due to the increasing number of applications targeted by the aforementioned actors. Today, social media have become an important part of our everyday life; hundreds of millions of users share a huge amount of data on such social media, which allow users to register through the creation of an account, to share content and to follow other members. Twitter is one of the most popular social media that allow users to post messages known as tweets. The emergence of Twitter has motivated a large number of research axes such as user account classification [Tavares, 13] [Daouadi, a19], topic classification [Khan, 17], event detection [Troudi, 18] and even terrorist detection [Rekik, 18]. This free micro-

blogging has been used not only by humans to create and share content, but also by bots (automated agents) to spread misinformation, to pollute content and even to perform terrorist propaganda [Nizzoli, 19] [Magdy, 16]. Classifying the patterns of user accounts into those of humans and those of bots can help users focus on valuable social accounts, get effective information, ensure their own security, avoid network traps etc. The challenge recently organized by the Defense Advanced Research Projects Agency (DARPA) for Twitter bot detection [Subrahmanian, 16] underscored the magnitude of the problem of bot detection. We define a human account as an account which has a non-automatic nature and is usually managed by a human. On the other hand, a bot account is an automatically managed profile (whose content is produced automatically). Current bot detection approaches are still not sufficient to detect bot activities accurately in term of the performance metrics. In fact, they have to face certain challenges owing to the problem inherent to text and the diversity of types of content propagated throughout Twitter. Moreover, messages from Twitter are imprecise, very short and written in an informal language full of spelling mistakes, abbreviations etc. In this paper, we will try to address the following research questions:

1. RQ1: Is it possible to accurately decide whether a user account belongs to a human or a bot using a minimal number of statistical features extracted from one single tweet?
2. RQ2: Is it possible to enrich previously-labeled datasets with more real examples of human and bot samples without the additional and very expensive annotation steps?

The rest of this paper is organized as follows: section 2 reviews the related works; section 3 presents our proposed framework; section 4 presents the results of our experiments; and finally, section 5 summarizes the main contributions of our proposed framework.

2 Related Works

Twitter user accounts have been analyzed across many dimensions. Several works have been done on a single and specific perspective such as those related to the bot detection ([Varol, 17], [Ferrara, 17], [Cresci, a17], [Bindu, 18], [Kudugunta, 18], [Morstatter, 16], [Lee, 11], [Wu,17], [Jain, 19], [Jain, 18], [Tavares, 17], [Singh, 18], [Stukal, 17], [Cresci, b17], [Daouadi, b19], [Cresci, 15], [Kantepe, 17], [Gilani, 17], and [Chen, 15]) and those related to the effects that bots have ([Cresci, 19], [Mazza, 19], [Gilani, 19] and [Cresci, 18]). The first part of this section discusses the ground truth acquisition methods for bot detection. The second part of this section reviews the literature of bot detection approaches. The third part of this section discusses our main contributions.

2.1 Ground Truth Acquisition Methods

Today, researchers through Twitter attempt to use three main ground truth acquisition methods for bot detection [Morstatter, 16]. The first one is the manual annotation method [Varol, 17], which relies on human annotators in the labeling phase. Although this method is a good technique and has been widely used, it has faced certain challenges owing to the time consumption, human expertise and the fact that it is not

scalable. The second method is known as suspended user lists method [Morstatter, 16]; it focuses on the social network's site itself in order to obtain the label. The researcher observes the users and tracks the ones that are removed or suspended by the site. This type of method is simple but has some drawbacks such as time consumption and the fact of being not scalable. The third commonly-used method is known as honeypots method [Lee, 11]; through this method, the researcher creates bot accounts in order to lure other bots. This allows collecting a large set of active bots with a high confidence. Although the existing ground truth acquisition methods are good techniques and have been heavily used by the related works, these methods are usually time-consuming and have a very expensive phase. To meet these challenges, we propose a new technique for enriching previous labeled datasets with more instances of real human and bot samples. The suggested technique is presented in the experimental section.

2.2 Bot Detection Approaches

Today, the problem of bot detection is one of the most critical challenges, especially on online social networks. According to [Ferrara, 16], Twitter bot detection approaches are grouped in three major types depending on the classification task. The first one is the crowdsourcing approach (also called blacklist approach [Wu, 17]), through this type of approach, expert workers are needed to detect bot accounts. The second one is known as graph-based approach. This type of approach is very expensive and more complex than other types of approaches. The third commonly used approach is known as features-based approach, different types of features are exploited to capture users' behaviors (e.g. content features, linguistic features, sentiment features, temporal features, post-frequency features, metadata of user profile features). According to the type of features used in the classification task, features-based approach include three main classes, namely content-based approaches, statistical-based approaches, and hybrid-based approaches.

2.2.1 Content-based Approaches

Through this type of approach, researchers attempted to use only the textual content parameters in order to perform the classification task. They used language-dependent features without taking into account the diversity of the types of content propagated throughout Twitter. In [Jain, 19], the authors proposed a semantic Long Short Term Memory in order to classify tweets according to their author type (spam and non-spam users). They used the textual content with Word2vec, Wordnet and Conceptnet in order to create a semantic word vector that has been used by Long Short Term Memory for the classification, this yielded F-measure result of 96.84%. In a similar [Jain, 18] but using Convolution Neural Network for the classification, this yielded F-measure result of 96.64%. In [Wu, 17], the authors used word2vec with traditional supervised learning algorithms, this yielded F-measure result between 92.83% and 94.25%.

2.2.2 Statistical-based Approaches

Through this type of approach, researchers attempted to use statistical parameters in order to perform the classification task. The main advantage of the statistical-based approach is that it uses language-independent features and avoids using the content parameters. In [Tavares, 17], the authors used parameters from the post frequency in

order to classify users as bots, humans or cyborgs, this yielded F-measure result of 88% using Random Forest algorithm. In [Cresci, a17] , the authors exploited digital DNA; they used the sequence of types of posts and the sequence of features-based content, and then they defined a similarity measure in order to classify users as bots or humans. This yielded F-measure result of 95.8% using an unsupervised approach. In [Ferrara, 17], the authors classified users as humans or bots; they used 10 parameters from the metadata of the users' profiles. This yielded AUC result of 92.0% using Random Forest algorithm. In [Daouadi, b19], the authors used statistical parameters extracted from the overall user's activity. The best experimental results are obtained with Deep Forest algorithm, this yielded Accuracy result of 97.55%.

2.2.3 Hybrid-based Approaches

Apart from the previously mentioned features-based approaches, the third commonly used approach is the hybrid-based approach. Through this type of approach, researchers attempted to combine the content-based approach and the statistical-based approach. In [Varol, 17], the authors used more than a thousand parameters; these included sentiment features, content features, user metadata features, friend features, network and timing features. This yielded AUC result of 95.0% using Random Forest algorithm. In [Singh, 18], the authors used 10 parameters from the content-based features, the user-based features, and the trust based features. This yielded Accuracy result of 92.1% using Random Forest algorithm. In [Stukal, 17], the authors used 42 parameters; these included metadata of user profile and tweeting features. This yielded F-measure result of 86,625% using an ensemble method of machine-learning algorithms. In [Lee, 11], the authors used 16 parameters from the user demographic features, the user-content features, the user-friendship networks and the user history features. This yielded AUC result of 98.4% using Random Forest algorithm. In [Kudugunta, 18], the authors proposed an approach known as contextual Long Short-Term Memory, which exploits both tweet content and metadata of user profile features. The best experimental results are obtained using Synthetic Minority Oversampling Technique, this yielded accuracy results >95%.

2.3 Motivation and Contribution

Although content-based and hybrid-based approaches are good techniques and were heavily used by the related works, these approaches were also time-consuming and they have faced the problems inherent to the text. On the other hand, statistical-based approaches have a lower computational cost and take a shorter time than the other types of approaches. In summary, current bot detection approaches are not sufficient to detect bot activities accurately in term of the performance metrics. We propose, Twitterbot+, a bot detection system that leveraged a minimal number of language-independent features extracted from one single tweet with temporal enrichment of a previously labeled datasets. We summarized our main contributions as the following:

- We designed a minimal number of statistical features that demonstrated their utility during this work. The suggested features were extracted from one single tweet of the user account, which makes Twitterbot+ suitable at tweet-level and user-level bot detection.

- We introduced a new technique for enriching existing labeled datasets by creating feature vectors from of the labeled user at different points of time. To the best of our knowledge, Twitterbot+ is the first bot detection model that enriches their training datasets from different points of time in the life cycle of the labeled user accounts. This is done without any additional and very expensive labeling steps.

3 Proposed Framework

Following the limits mentioned in our related works, we propose Twitterbot+ for accurate and real-time bot detection from Twitter. In general, the task of supervised bot detection begins with a training set $U = (U_1 \ldots U_N)$ of users that are labeled with a class $c_i \in C$ (ex. bot or human). Then, the task is to find the optimal features set $F = (f_1 \ldots f_n)$ that is able to correctly assign a class (c_i) to a new user (U_i) that has the feature vector $F^i = (f_1^i \ldots f_n^i)$. Our proposition consists of three main steps, namely data collection, features extraction and classification.

3.1 Data Collection

In this step, our main aim was to use the APIs offered by Twitter. We used Twitter Timeline API, which allows collecting the most recent post of a given user. However, we rely on three public datasets which were published in [Lee, 11], [Cresci, b17] and [Cresci, 15]. These datasets contain a user_ids with a label as Human or Bot. We used the user_id with the Twitter Timeline (API) in order to collect the most recent post of each labeled user. The post comes with an information from of Tweet, Time and User attributes. The Tweet attribute contains the content and the statistics of the tweet. The Time attribute contains the time and the date of the tweet. The User attribute contains information about the user's profile. We observed the users' accounts' statuses via the statuses/user_Timeline API endpoint. These statuses can take one of four values: active account, suspended account, removed account and private account. However, we could not access the post of user accounts that were suspended, private or removed. In this study, we used only the active accounts. Each recent post (P_i) of each active labeled user (U_i) is retrieved, and this way the next step could follow.

3.2 Feature Extraction

Although many proposed approaches use a very large set of features (e.g. [Varol, 17] uses over 1000 features), a recent study [Ferrara, 17] shows that a limited number of features are enough in order to obtain similar performance measures. The choice of using a size limit of features is motivated by both model efficiency and interpretability reasons. Based on the existing works, we designed a set of minimal numbers of statistical features that were extracted from of the User attribute (described in the previous subsection) and demonstrated their utility during this work. The suggested features are:

1. Average number of tweets per day.
2. Status count (the total number of tweets issued by the user).
3. Location (whether the user has provided his/her location).

4. Description (whether a description is provided in association with the user profile).
5. Follower count (the number of followers of the user).
6. Following count (the number of users that this user follows).
7. The ratio of Follower count to Following count.
8. URL (whether a URL is provided in association with his/her profile).
9. Favorite's counts (the number of favorites done by the user).
10. Verified (whether the user has a certified account).
11. Default profile (whether the user profile has the default settings).
12. Lists count (the number of public lists that this user is a member of).
13. Age (the age of the user account in days).
14. Geo (whether the user has enabled the possibility of geo-tagging his/her tweets).
15. Default image background (whether the user has uploaded his/her own image).

From each collected (P_i), we created a feature vector $F^i = (f_1^i \ldots f_n^i)$, where n=15 is the number of the proposed features.

3.3 Classification

In this step, our main objective was to use a supervised learning approach in order to classify the feature vectors $F^i = (f_1^i \ldots f_n^i)$ that has $c_i \in C$ (Human or Bot). First, we chose the algorithm that performs better with our proposed features. Several supervised learning algorithms were tested; these included AdaBoost (AB), Bagging (B), Random Forest (RF), Simple Logistic (SL) and LogitBoost (LB). These algorithms were used with their standard parameters in order to ensure the comparability between them. Second, we enriched our interested datasets by creating feature vectors form each labeled user account at different time points. However, each active user U_i will have $\{F_{t1}^i \ldots F_{tM}^i\}$ feature vectors where M is the number of time points. Several experiments were performed in order to compare our proposed temporal enrichment of a previously labeled user accounts against the baseline. However, we built three models, one containing all the datasets, one containing all the datasets enriched by the Synthetic Minority Oversampling Technique (SMOTE) [Chawla, 02] (as suggested by [Kudugunta, 18]) and one containing all the datasets in more than one point in time.

4 Experimental Settings and Results

Our proposition is evaluated on three benchmark datasets, these were published by [Lee, 11], [Cresci, b17] and [Cresci, 15]. Table 1 shows an overview of our interested datasets. The first one presents a combination of content-polluter accounts and legitimate accounts; we refer to this one as D1. The second one presents a combination of genuine accounts, social spambots #1, social spambots #2 and social spambots #3

accounts; we refer to this as D2. The third one presents a combination of genuine accounts and fake accounts; we refer to this as D3.

Label	Human	Bot	Active Human	Active Bot
D1	19276	22223	13472	14713
D2	3474	4912	2718	4293
D3	1950	3351	1665	715
D1+D2+D3	24700	30486	17855	19721

Table 1: Overview of our interested ground truth.

To evaluate the performance of Twitterbot+, we are interested in several metrics. We used the matrix confusion, which allows the visualization of the performance of a classification algorithm. We used the corresponding parameters; in our case, they are called:

- True Bot (TB): the instances correctly classified as bot.
- True Human (TH): the instances correctly classified as human.
- False Bot (FB): the instances incorrectly classified as bot.
- False Human (FH): the instances incorrectly classified as human.

Using the aforementioned parameters, the performance metrics that can be drawn are Precision (P), Recall (R), F1-measure (F1) and Matthews Correlation Coefficient (MCC). These metrics are calculated based on the following Equations 1,2,3 and 4:

$$P = TB/(TB + FB) \quad (1)$$

$$R = TB/(TB + FH) \quad (2)$$

$$F1 = 2((P * R)/(P + R)) \quad (3)$$

$$MCC = TB * TH - FB * FH/\sqrt{(TB + FH)(FB + TH)(TB + FB)(TH + FH)} \quad (4)$$

We also used the Area Under the Curve (AUC or AUC/ROC) measure. This measure "is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance" [Fawcett, 06] (in our case, positive instance is a bot instance and negative instance is a human instance). The 10-fold cross validation approach was performed in order to calculate the performance metrics. Each dataset was split into 10 equally sized groups while preserving a balance of the bot and human samples of the original datasets. One of these groups was used for testing whereas the remainder were used for training. This procedure was iterated 10 times and the averaging of the performance metrics scores was obtained across the ten iterations of cross validation.

The first set of our experiments evaluate the performance metrics of our proposed features with previously mentioned supervised learning algorithms. As can be concluded from the experiment presented in Table 2, the RF algorithm outperformed the other algorithms in the corresponding models:

- D1: this yielded (P=90.6%, R=90.6%, F1=90.6%, AUC=96.6%).
- D2: this yielded (P=98.6%, R=98.6%, F1=98.6%, AUC=99.6%).
- D3: this yielded (P=93.0 %, R=93.1%, F1=93.0%, AUC=97.2%).
- D1+D2+D3: this yielded (P=91.4 %, R=91.4%, F1=91.4 %, AUC=97.0%).

Algorithm	***Metric %***	***D1***	***D2***	***D3***	***D1+D2+D3***
AB	*P*	*87.2*	*96.9*	*92.3*	*85.3*
	R	*86.4*	*96.9*	*92.4*	*85.3*
	F1	*86.4*	*96.9*	*92.3*	*85.3*
	AUC	*92.3*	*99.3*	*96.3*	*92.9*
B	*P*	*90.4*	*98.3*	*92.2*	*91.0*
	R	*90.3*	*98.3*	*92.2*	*91.0*
	F1	*90.3*	*98.3*	*92.1*	*91.0*
	AUC	*96.4*	*99.5*	*97.0*	*96.8*
RF	*P*	***90.6***	***98.6***	***93.0***	***91.4***
	R	***90.6***	***98.6***	***93.1***	***91.4***
	F1	***90.6***	***98.6***	***93.0***	***91.4***
	AUC	***96.6***	***99.6***	***97.2***	***97.0***
SL	*P*	*79.9*	*95.9*	*92.0*	*77.9*
	R	*79.9*	*95.9*	*92.1*	*77.9*
	F1	*79.9*	*95.9*	*92.0*	*77.8*
	AUC	*87.6*	*98.8*	*95.0*	*86.1*
LB	*P*	*87.5*	*97.7*	*91.2*	*86.9*
	R	*87.2*	*97.7*	*91.2*	*86.9*
	F1	*87.3*	*97.7*	*91.2*	*86.9*
	AUC	*94.4*	*99.4*	*94.7*	*93.8*

Table 2: Performance comparison of different supervised learning algorithms.

The second set of our experiments compare Twitterbot+ against the baselines. However, we are interested on the works published by [Ferrara, 17] and [Kudugunta, 18]. From the experiment presented in Table 3, Twitterbot+ yielded F1 result of 91.4% (Vs. 88.3% [Ferrara, 17] and 90.1% [Kudugunta, 18]), AUC result of 97.0% (Vs. 94.6% [Ferrara, 17] and 95.8% [Kudugunta, 18]) and MCC result of 82.7% (Vs. 76.5% [Ferrara, 17] and 78.6% [Kudugunta, 18]).

Work	**F1%**	**AUC %**	**MCC %**
[Ferrara, 17]	88.3	94.6	76.5
[Kudugunta, 18]	90.1	95.8	78.6
Twitterbot+	**91.4**	**97.0**	**82.7**

Table 3: Performance comparison of our proposed features with baselines.

The third set of our experiments answer to the second research question, we built on the hypothesis that capturing the behaviors of the labeled user types from different

points of time can maximize the accuracy results while enriching previously labeled datasets by real samples of both human and bot. First, the majority, if not all of previously mentioned related works use information from a single observation from the labeled accounts (i.e. creating feature vector of the labeled datasets from one time point). In fact, a single observation from the user accounts is not enough to predict the type of user account accurately; this is due to the fact that both types change their behaviors during their life cycle. Second, the majority, if not all of the related works enrich existing labeled datasets either by using the additional and very expensive labeling steps [Varol, 17] or by creating new random samples based on the original feature vectors [Kudugunta, 18]. Thus, temporal enrichment seems to be an interesting solution for enriching previously labeled datasets with more real examples of human and bot. To validate our hypothesis, we created feature vectors from previously labeled datasets at t=0, t=10 days and t=100 days. We performed several experiments over the SMOTE algorithm baseline and our proposed temporal enrichment. As can be concluded from the experiment presented in Table 4, our proposed temporal enrichment achieves higher accuracy results than the baseline. This yielded F1 result of 97.9% (Vs. 93.4% [Kudugunta, 18]), AUC result of 99.7% (Vs. 98.3% [Kudugunta, 18]) and MCC result of 95.8% (Vs. 86.8% [Kudugunta, 18]).

Datasets	**F1%**	**AUC %**	**MCC %**
Original	91.4	97.0	82.7
Original + SMOTE [Kudugunta, 18]	93.4	98.3	86.8
Original + temporal enrichment (Ours)	**97.9**	**99.7**	**95.8**

Table 4: Performance comparison of our proposed technique for enriching previous labeled datasets and SMOTE baseline.

5 Conclusion and Future Prospects

The approach taken in Twitterbot+ uses a minimal number of statistical features extracted from one single tweet with temporal enrichment of a previously labeled datasets, yet it yields high accuracy results (>95%), which is the lack in the majority of the state-of-the-art approaches. In addition, Twitterbot+ is not confronted by the problems inherent to the text. Moreover, Twitterbot+ is the first bot detection model that enriches their training data with temporal enrichment from the previously labeled user accounts. The evaluation of Twitterbot+ was done using three benchmark datasets. This will allow researchers to compare different systems fairly against our system. Experiment shows that simply using minimal parameters from of metadata of user accounts is sufficient to detect bot account accurately and quickly. As future works, we will extend our work to be able to detect the topic and the sentiment of the classified tweet. From a research standpoint, we will use Twitterbot+ to analyze Twitter conversations in different contexts to determine the extent of the interaction of humans and bots with public discourse, and to understand how their sophistication and capabilities evolve over time.

References

[Bindu, 18] Bindu, P. V., Mishra, R., Thilagam, P. S.: Discovering spammer communities in Twitter. Journal of Intelligent Information Systems, 51(3), 503-527, (2018).

[Chawla, 02] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P.: SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357, (2002).

[Chen, 15] Chen, C., Zhang, J., Xie, Y., Xiang, Y., Zhou, W., Hassan, M. M., Alrubaian, M.: A performance evaluation of machine learning-based streaming spam tweets detection. IEEE Transactions on Computational social systems, 2(3), 65-76, (2015).

[Cresci, 15] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: Fame for sale: Efficient detection of fake Twitter followers. Decision Support Systems, 80, 56-71, (2015).

[Cresci, a17] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: Social fingerprinting detection of spambot groups through DNA-inspired behavioral modeling. IEEE Transactions on Dependable and Secure Computing, 15(4), 561-576, (2017).

[Cresci, b17] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In Proc. Int. Conf on World Wide Web companion (pp. 963-972), Australia, (2017, April).

[Cresci, 18] Cresci, S., Lillo, F., Regoli, D., Tardelli, S., Tesconi, M.: $ FAKE: Evidence of Spam and Bot Activity in Stock Microblogs on Twitter. In Proc. Int. Conf on Web and Social Media, (pp.580-583), (2018, June).

[Cresci, 19] Cresci, S., Lillo, F., Regoli, D., Tardelli, S., Tesconi, M.: Cashtag piggybacking: Uncovering spam and bot activity in stock microblogs on Twitter. ACM Transactions on the Web, 13(2), 11, (2019).

[Daouadi, a19] Daouadi, K. E., Zghal Rebaï, R., Amous, I.: Organization, Bot, or Human: Towards an Efficient Twitter User Classification. Computación y Sistemas, 23(2), 273-280, (2019).

[Daouadi, b19] Daouadi, K. E., Rebaï, R. Z., Amous, I. Bot Detection on Online Social Networks Using Deep Forest. In Proc. Int. Conf on Computer Science On-line Conference (pp. 307-315), (2019, April).

[Fawcett, 06] Fawcett, T.: An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874, (2006).

[Ferrara, 16] Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Communications of the ACM, 59(7), 96-104, (2016).

[Ferrara, 17] Ferrara, E.: Disinformation and social bot operations in the run up to the 2017 French presidential election. First Monday, 22(8), (2017).

[Gilani, 17] Gilani, Z., Kochmar, E., Crowcroft, J.: Classification of twitter accounts into automated agents and human users. In Proc. Int. Conf on Advances in Social Networks Analysis and Mining (pp. 489-496), (2017, July).

[Gilani, 19] Gilani, Z., Farahbakhsh, R., Tyson, G., Crowcroft, J.: A Large-scale Behavioural Analysis of Bots and Humans on Twitter. ACM Transactions on the Web, 13(1), 7, (2019).

[Jain, 18] Jain, G., Sharma, M., Agarwal, B.: Spam detection on social media using semantic convolutional neural network. International Journal of Knowledge Discovery in Bioinformatics, 8(1), 12-26, (2018).

[Jain, 19] Jain, G., Sharma, M., Agarwal, B.: Optimizing semantic LSTM for spam detection. International Journal of Information Technology, 11(2), 239-250, (2019).

[Kantepe, 17] Kantepe, M., Ganiz, M. C.: Preprocessing framework for twitter bot detection. In Proc. Int. Conf on Computer Science and Engineering (pp. 630-634), (2017, October).

[Khan, 17] Khan, I., Naqvi, S. K., Alam, M., & Rizvi, S. N. A.: An efficient framework for real-time tweet classification. International Journal of Information Technology, 9(2), 215-221, (2017).

[Kudugunta, 18] Kudugunta, S., Ferrara, E.: Deep neural networks for bot detection. Information Sciences, 467, 312-322, (2018).

[Lee, 11] Lee, K., Eoff, B. D., Caverlee, J.: Seven months with the devils: A long-term study of content polluters on twitter. In Proc. Int. Conf AAAI on Weblogs and Social Media, Spain, (2011, July).

[Magdy, 16] Magdy, W., Darwish, K., Weber, I.: # FailedRevolutions: Using Twitter to Study the Antecedents of ISIS Support. In AAAI Spring Symposium Series. (2016, March).

[Mazza, 19] Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., Tesconi, M.: RTbust: Exploiting Temporal Patterns for Botnet Detection on Twitter. In Proc. Int. Conf on Web Science, (pp.183-192), (2019).

[Morstatter, 16] Morstatter, F., Wu, L., Nazer, T. H., Carley, K. M., Liu, H.: A new approach to bot detection: striking the balance between precision and recall. In Proc. Int. Conf on Advances in Social Networks Analysis and Mining (pp. 533-540), USA, (2016, August).

[Nizzoli, 19] Nizzoli, L., Avvenuti, M., Cresci, S., Tesconi, M.: Extremist Propaganda Tweet Classification with Deep Learning in Realistic Scenarios. In Proc. Int. Conf on Web Science (pp. 203-204), (2019, June).

[Rekik, 18] Rekik, A., Ameur, H., Abid, A., Mbarek, A., Kardamine, W., Jamoussi, S., Hamadou, A. B.: Building an Arabic Social Corpus for Dangerous Profile Extraction on Social Networks. Computación y Sistemas, 22(4), (2018).

[Singh, 18] Singh, M., Bansal, D., Sofat, S.: Who is who on twitter–spammer, fake or compromised account? A tool to reveal true identity in real-time. Cybernetics and Systems, 49(1), 1-25, (2018).

[Stukal, 17] Stukal, D., Sanovich, S., Bonneau, R., Tucker, J. A.: Detecting bots on Russian political Twitter. Big data, 5(4), 310-324, (2017).

[Subrahmanian, 16] Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Menczer, F.: The DARPA Twitter bot challenge. Computer, 49(6), 38-46, (2016).

[Tavares, 13] Tavares, G., Faisal, A.: Scaling-laws of human broadcast communication enable distinction between human, corporate and robot twitter users. PloS one, 8(7), e65774, (2013).

[Tavares, 17] Tavares, G., Mastelini, S.: User classification on online social networks by post frequency. In Anais do XIII Simpósio Brasileiro de Sistemas de Informação (pp. 464-471). Brazil, (2017, May).

[Troudi, 18] Troudi, A., Zayani, C. A., Jamoussi, S., Amor, I. A. B.: A new mashup based method for event detection from social media. Information Systems Frontiers, 20(5), 981-992, (2018).

[Varol, 17] Varol, O., Ferrara, E., Davis, C. A., Menczer, F., Flammini, A.: Online human-bot interactions: Detection, estimation, and characterization. In Proc. Int. Conf on web and social media. (2017, May). https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587.

[Wu, 17] Wu, T., Wen, S., Liu, S., Zhang, J., Xiang, Y., Alrubaian, M., Hassan, M. M.: Detecting spamming activities in twitter based on deep-learning technique. Concurrency and Computation: Practice and Experience, 29(19), e4209, (2017).