Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 46 (2025), No. 6, pp. 1851–1859
DOI : 10.47974/JIOS-2010

Design of an efficient deep pattern analysis model for the identification of bot-based fake profiles

Bhrugumalla L. V. S. Aditya *
Sachi Nandan Mohanty †
School of Computer Science and Engineering
VIT-AP University
Amaravati 522237
Andhra Pradesh
India

Abstract

To ensure the integrity and safety of online interactions, many fake profiles on social media are often created by harmful bots. Despite the tactics already employed, their efficacy is frequently compromised by flaws like the reliance on static profile traits and the inability to consider changing behavioral patterns. These flaws call for a more reliable and flexible method to identify bot-based fake profiles precisely. This paper offers a novel model that uses Deep Dyna-Q Networks (DDQN) to fill this gap, improving the precision and speed of bogus profile identification. In contrast to conventional approaches, the proposed model takes a closer look at an account's activity, followers, and following in addition to its profile traits. This method considers the number and quality of these parameters and, as a result, can spot subtle patterns that may point to a fake account for real-time scenarios. The suggested methodology uses multidomain feature analysis, making understanding intricate, linked behavioral patterns easier. This in-depth investigation makes it easier to spot the abnormalities in activity, correlations in followers and followings, and discrepancies in profile information that are direct signs of fake profiles created by bots. The study's findings show that the suggested model performs better than others. In addition to a decrease in detection delay, there is an improvement in precision, accuracy, recall, A.U.C., sensitivity, and specificity levels. These enhancements demonstrate the model's potency and highlight the

* *ORCID-ID:* 0000-0002-5988-0757
† *ORCID-ID:* 0000-0002-4939-0797

* *E-mail:* blvsaditya@gmail.com (Corresponding Author)
† *E-mail:* sachinandan09@gmail.com

possibility of using cutting-edge deep learning algorithms to address the growing problem of fake profiles on social media sites.

Subject Classification: *68T05, 68T09, 68T27, 68R10.*

Keywords: *Bot-generated profiles, Deep dyna-Q networks, Fake profile detection, Multidomain feature analysis.*

1. Introduction

Due to the digital age's communication revolution, social media platforms are crucial for interactions, information sharing, and networking. Bots generating fake profiles to deceive or influence public opinion are problematic online. Bot accounts generate 29% of Internet traffic, according to Oxford University. Many methods fail to eliminate fake profiles. Many approaches use static profile elements like usernames, photographs, and location. These signs are easily fabricated and don't verify the report. Traditional detection methods neglect dynamic behavioural patterns, allowing intricate bot profiles to go unnoticed. Scalability and timing concerns plague current systems. More social media users mean more data to evaluate, slowing discovery and raising the risk of hostile false personas. Therefore, a more reliable, scalable, and dynamic detection system is needed to examine complicated patterns and respond to bot activity [1]. Bot-based false profile identification using Deep Dyna-Q Networks and multidomain feature analysis is presented in this paper. Static profile information, activity, followers, and following are detailed by the model. It explains suspicious conduct including restricted or repetitive activity and irregular followers and followings. This paradigm is flexible and scalable. Pattern recognition helps it adapt to bot algorithm changes. Model handles enormous data with multidomain feature analysis for quick and successful identification [2].

2. Review of existing models used for identification of fake profiles

Fake bot profiles are hard to spot. Academics and industry professionals have produced many detection models due to its popularity. These models define user behavior, account properties, and network links using GAN and LSTM feature extraction [3].

The simplest DeepSBD-type identification uses username, photo, and location [4]. This method is useful for catching simple bots but has limitations. The 'A Star Model' (A.S.M. for different circumstances) shows

bot builders how to simply alter these qualities to make their creations feel real [5]. SVMs, Random Forests, and Decision Trees detect fake profiles. These models incorporate behavioral and extractive accounts [6]. Popular methods scale poorly, especially with social media data. Static data training makes these models bot-unfriendly. Some algorithms detect suspicious account follower behavior using network data [7]. This works because bot accounts follow each other or have a high follower-to-follower ratio [8]. Network structure concentration makes these models neglect subtler behaviors and patterns [9]. CNN-based deep learning is promising [10]. Complex patterns from unstructured data can help these computers detect increasingly complex bot activities. They are effective but computationally expensive and require many labeled training data samples [11]. Graph-based models use structural data from graph embeddings and social networks [12].

3. Proposed design of an efficient deep pattern analysis model for identification of bot-based fake profiles

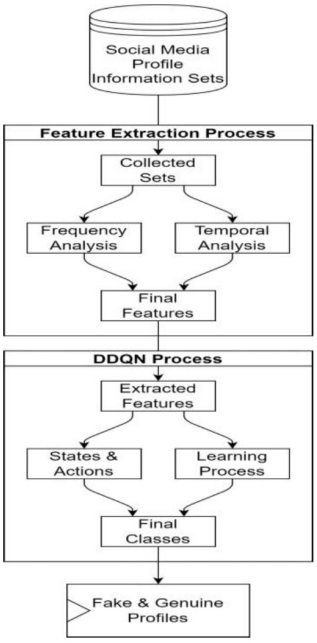

Figure 1

Proposed Model for Identification of Bot-based Fake Profiles

The frequency of posts per unit of time (e.g., per day) can be calculated as the ratio of the total number of posts to the time elapsed between the first and last posts via equation 1. Also, Figure 1 represents the bot-based fake profile.

$$Frequency = \frac{P}{T[n] - T[1]} \tag{1}$$

Furthermore, the variability in the time intervals between consecutive posts can be assessed using the standard deviation of the time differences via equation 2,

$$Variability = SD(T[2:n] - T[1:n-1]) \tag{2}$$

Where, *SD* represents the Standard Deviation of the input variable and is estimated via equation 3,

$$SD = \sqrt{\frac{\sum_{i=1}^{N}\left(x(i) - \sum_{j=1}^{N}\frac{x(j)}{N}\right)^2}{N}} \tag{3}$$

the similarity between the post content and a collection of pre-defined generic or promotional templates (*generic templates*) was measured using cosine similarity via equation 4,

$$Similarity = \frac{T \ .GT}{||T|| * ||GT||} \tag{4}$$

The engagement ratio was computed as the ratio of interactions to the number of followers via equation 5,

$$Engagement\ Ratio = \frac{I}{F} \tag{5}$$

To assess the timing of posts, we consider the number of posts made during peak hours (Ppeak) and off-peak hours (Poffpeak) relative to the total number of posts (P total). The proportion of posts made during peak hours and off-peak hours was estimated via equations 6 & 7 as follows,

$$PHP = \frac{Ppeak}{Ptotal} \tag{6}$$

$$OHP = \frac{Poffpeak}{Ptotal} \tag{7}$$

Given an augmented set of profile feature sets in the form of 1D signal x(t), the continuous Fourier Transform is defined via equation 8,

$$X(f) = \int x(t) * e^{-2\pi jft} dt \tag{8}$$

Given an augmented set of profile feature sets in the form of 1D signal x(t), the continuous Wavelet Transform was estimated via equation 9,

$$W(a,b) = \int x(t) * \psi * \left(\frac{t-b}{a} \right) dt \tag{9}$$

The Short-Time Fourier Transform allows for time-varying frequency analysis of profile feature sets in the form of 1D signals. It segments the signal into short windows, and the discrete STFT was evaluated via equation 10,

$$STFT(n,m) = \sum x(k) * w(k-n) * e^{-\frac{2\pi jmk}{N}} \tag{10}$$

The Discrete Wavelet Transform decomposes profile feature sets into 1D signals into different frequency bands at various scales. It is typically implemented using a filter bank approach. Given a discrete signal x[n], the DWT was computed using wavelet filters h[n] and g[n] via equation 11,

$$DWT[i,n] = \sum x[k] * h[k - 2^n * n] \tag{11}$$

DWT [i, n] represents the DWT coefficients at level I, and h[k] is the low-pass filter. The Singular Value Decomposition was used for dimensionality reduction and data compression operations. It was estimated via equation 12,

$$A = U * \Sigma * V^T \tag{12}$$

. After this, the Z-Transform was estimated via equation 13,

$$X(z) = \sum x[n] * z^{-n} \tag{13}$$

The Hilbert Transform can be applied to profile feature sets in the form of 1D signals to obtain the analytic representation. It provides amplitude and phase information, which is estimated via equation 14,

$$H(x(t)) = \left(\frac{1}{\pi} \right) * \frac{\int x(t')}{t-t'} dt' \tag{14}$$

The Discrete Cosine Transform was used to estimate the Entropy of the profile parameters via equation 15,

$$X(k) = \sum x[n] * cos\left(\frac{\pi}{N} * (n + 0.5) * k\right) \tag{15}$$

The Mellin Transform was used to provide features for pattern analysis via equation 16,

$$M\{x(t)\} = \int t^{s-1} * x(t) dt \tag{16}$$

The Convolution Operations convert BPFV into an augmented set of Convolutional Features (C.F.) via equation 17,

$$CF = \sum_{a=-\frac{m}{2}}^{\frac{m}{2}} BPFV(i-a) * LReLU\left(\frac{m+2a}{2}\right) \tag{17}$$

Where, m & a are the size of window & stride kernels, *LReLU* is used for feature activation, and is represented via equation 18,

$$LReLU(y) = \max(c * y, y) \tag{18}$$

Where c is an iterative constant, which is used to activate these features. The output from these operations is flattened into a 1D vector, which allows it to be used as input to the Q Network process. The flatten operation converts the 2D feature maps into a 1D vector via equation 19,

$$flatten(x) = [x[1,1], x[1,2], \ldots, x[1, width],$$

$$x[2,1], x[2,2], \ldots,$$

$$x[r, c]] \tag{19}$$

The Q Value for each profile type was expressed via equations 20 & 21 as follows,

$$Q(s, fake') = NNfake(xflatten) \tag{20}$$

$$Q(s, genuine') = NNgenuine(xflatten) \tag{21}$$

Where, $NNfake(xflatten)$ and $NNgenuine(xflatten)$ represent the outputs of the neural networks for the 'fake' and 'genuine' profile types, respectively, which are estimated via equation 22,

$$NN(out) = SoftMax\left(\sum_{i=1}^{N_f} f(i) * w(i) + b(i)\right) \tag{23}$$

Where, $SoftMax$ represents SoftMax-based activation function, while w & b means the weights & biases for the Convolution Process.

The Q Value update equation for each profile type via equation 23,

$$Q(s,action) = Q(s,action) + \alpha * [R + \gamma * max\{a'\}Q(s',a') - Q(s,action)] \quad (23)$$

Results

	Accuracy	Precision	Recall	Delay(ms)
[2]	82.80	79.21	82.48	159.48
[3]	87.39	82.74	85.04	164.63
[4]	92.00	86.15	89.50	156.77
This model	97.76	94.02	94.34	116.78

5. Conclusion

This study sought a powerful deep pattern analysis algorithm to detect bot-created social media profiles. The recommended model faced deep SBD, GAN, and A.S.M. Precision, accuracy, recall, A.U.C., and specificity were assessed throughout NTS sizes. The extended research demonstrates that the proposed technique handles bot-created fake profiles better. The recommended model outperformed alternatives in precision, accuracy, recall, A.U.C., and specificity across all dataset sizes. The approach is reliable and can detect bot-based bogus profiles. Better precision, accuracy, and memory allow the proposed method to detect minor patterns and anomalies in fake profiles. Deep Dyna-Q Networks (DDQN) detect complex and dynamic fraudulent profiles in real time using multimodal feature analysis and bot behavior analysis. The model decreased processor latency. The model may reject bot-created profiles and intervene in real time with fast categorization.

High A.U.C. and specificity values show that the proposed model can discriminate bot-created profiles from accurate profiles, eliminating false positives and negatives. Online security and bot and misinformation prevention may improve. Bot-based phony profiles may benefit from DDQN, multimodal feature, and bot analysis. Higher performance, accuracy, and processing speed make the model reliable and versatile for reducing the harm of fake social media profiles. This study proposes a

practical and accurate method to identify and eliminate bot-generated fake profiles for online trust and authenticity.

References

[1] J.-M. Chen, "Multi-vendor and multi-buyer collaborative planning, forecasting and replenishment model," *Journal of Statistics and Management Systems,* vol. 28, no. 6, pp. 1023–1036 (2025).

[2] G. B. Mohammad, S. Shitharth, and P. R. Kumar, "Integrated machine learning model for an URL phishing detection," *International Journal of Grid and Distributed Computing,* vol. 14, no. 1, pp. 513–529 (2021).

[3] H. Garg, S. Jhunthra, R. Goel, S. Sharma, V. Gupta, R. Sharma, and P. Dass, "Understanding user polarisation regarding COVID-19 vaccines through social network analysis," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 27, no. 8, pp. 2301–2336 (2024).

[4] M. Fazil, A. K. Sah, and M. Abulaish, "DeepSBD: A deep neural network model with attention mechanism for social bot detection," *IEEE Transactions on Information Forensics and Security,* vol. 16, pp. 4211–4223 (2021).

[5] Muskan and S. S. C. (Gonder), "A theoretical approach to find the solution of population based mathematical model," *Journal of Interdisciplinary Mathematics,* vol. 27, no. 8, pp. 1735–1740 (2024).

[6] B. Shilpa, P. R. Kumar, and R. K. Jha, "Spreading factor optimization for interference mitigation in dense indoor LoRa networks," in Proc. 2023 IEEE IAS Global Conf. Emerging Technologies (GlobConET), pp. 1–5 (2023).

[7] P. R. Kumar, "Wireless mobile charger using inductive coupling," *Journal of Emerging Technologies and Innovative Research* (JETIR), vol. 5, no. 10, pp. 40–44 (2018).

[8] S. Khan, et al., "HCovBi-Caps: Hate speech detection using convolutional and bi-directional gated recurrent unit with capsule network," *IEEE Access*, vol. 10, pp. 7881–7894 (2022).

[9] M. Aljabri, R. Zagrouba, A. Shaahid, et al., "Machine learning-based social media bot detection: A comprehensive literature review," *Social Network Analysis and Mining*, vol. 13, no. 20 (2023).

[10] P. R. Kumar and B. Shilpa, "An IoT-based smart healthcare system with edge intelligence computing," in Reconnoitering the Landscape

of Edge Intelligence in Healthcare, Apple Academic Press, pp. 31–46 (2024).

[11] T. He and N. Lia, "Age-structured model with varied contact patterns and vaccination of COVID-19 in Liaoning Province, China," *Journal of Statistics and Management Systems*, vol. 27, no. 8, pp. 1525–1539 (2024).

[12] A. Aldayel and W. Magdy, "Characterizing the role of bots in polarized stance on social media," *Social Network Analysis and Mining,* vol. 12, no. 30 (2022).

Received December, 2024