Supervised Learning vs Reinforcement Learning Models for Fake News Detection

Alexander Wang & Hao-Lun Hsu

Received May 02, 2024
Accepted June 26, 2024
Electronic access July 15, 2024

This research explores the effectiveness of supervised learning and reinforcement learning models in detecting fake news, which is a growing concern in the digital age. We compare the performance of these two machine-learning approaches using a dataset containing both real and fake news articles. For the supervised learning approach, we trained a neural network using a diverse dataset of labeled news articles. On the other hand, the reinforcement learning approach involved training an agent using a reward-based system. The agent learned to identify fake news by interacting with human responses to fake news. The results showed that the supervised learning models achieved a high accuracy rate of up to 95% on the test set, indicating their strong capability to recognize patterns indicative of fake news. In contrast, the reinforcement learning model only achieves an average reward of 10.8 out of a theoretical maximum of 20. This result indicates that the reinforcement learning model is only slightly better than random chance at correctly identifying fake news or real news, highlighting the need to consider the main text as part of the input. We discuss the implications of these findings for developing more robust and dynamic fake news detection systems, highlighting potential areas for future research.

Keywords: Reinforcement Learning, Supervised Learning, Machine Learning, Fake News Detection, Neural Networks

Introduction

Background and Context

In an era marked by the burgeoning popularity of social media, the ease of message dissemination has reached unprecedented levels. However, this proliferation has engendered a concurrent surge in the propagation of misinformation, a phenomenon exacerbated during periods of crisis. A stark illustration of this phenomenon emerged during the 2020 COVID-19 pandemic, wherein numerous individuals faced hospitalization consequent to the misguided belief that bleach could cure the virus[1]. Fortunately, the evolution of artificial intelligence (AI) and machine learning (ML) furnishes us with potent tools to counteract misinformation and combat the spread of fake news.

The emergence of fake news is a pervasive challenge in the contemporary digital landscape. Defined as the dissemination of false information masquerading as factual, fake news permeates many media platforms, including news outlets and social networking sites. Pressured to captivate audiences, news agencies often prioritize sensational headlines, amplifying the risk of misinformation. Similarly, the decentralized nature of social media platforms such as X (Twitter), Reddit, and Quora facilitates the unrestricted dissemination of content, impeding the verification of authenticity. Consequently, as the user base expands, the diffusion of fake news becomes increasingly difficult. Thus, developing robust tools to combat the spread of misinformation becomes increasingly imperative.

ML, a branch of AI, embodies a transformative paradigm wherein algorithms and statistical models facilitate autonomous learning and performance enhancement from data sans explicit programming. In essence, ML empowers machines to discern patterns, make predictions, and derive decisions predicated on data analysis. The significance of ML in contemporary society is manifold, underpinning enhanced data-driven decision-making, automation, and operational efficiency. Within the realm of ML, we mainly investigate supervised learning[2,3] and reinforcement learning[4] methodologies for the detection of fake news.

Pior works have endeavored to harness ML methodologies to solve fake news problems, albeit adopting different approaches. For instance, Khanam et al. employed traditional ML algorithms such as random forest (RF), support vector machine (SVM), and k-nearest neighbors (KNN), with XGBoost attaining peak accuracy exceeding 75%, supplemented by the integration of additional textual analysis features to augment precision outcomes[2]. On the other hand, other researchers have explored reinforcement learning techniques to address this challenge. For example, Mosallanezhad et al. devised the REinforced Adaptive Learning Fake News Detection (REAL-FND) model, leveraging reinforcement learning principles[4]. In addition, Wang et al. devised a multifaceted system comprising an annotator, a reinforced selector, and a core fake news detection model to segregate falsehoods from genuine information[5].

0.1 Problem Statement and Methodology Overview

In this work, we use ML techniques to tackle the challenge of discerning the authenticity of textual content sourced from Twitter, now known as X, distinguishing between authentic information and false narratives. The inquiry centers on two primary questions: Firstly, can ML methodologies directly determine the authenticity of Twitter text (hereafter referred to as X) by categorizing it as real or fake news? Secondly, can ML indirectly predict the authenticity of the same textual data by approximating human responses?

We aim to explore a novel ML framework that integrates approximated human responses, aiming to ascertain the authenticity of textual data without direct analysis of the text content. This experimental approach sought to evaluate the feasibility of predicting the veracity of text solely based on human evaluative cues. This methodology will be compared with a conventional ML model tasked with directly assessing the authenticity of X text by analyzing its textual content. Our final goal is to increase the prediction accuracy and performance of both models.

We focus on using a supervised learning neural network to identify fake news according to the corresponding main texts (human response) and reinforcement learning with a customized-built gym environment to approximate human response with online interaction in simulation for this problem. We chose these models as compared to other models such as RF and SVM, because from our research, there has not been much research done using these models on identifying fake news.

To evaluate our proposed method, we mainly deployed our approaches on a dataset retrieved from Kaggle[6]. More information regarding the dataset is introduced in the materials and methods section.

Preliminary

Supervised learning is a method in ML that involves training a model to make predictions or classifications based on a labeled dataset. In this context, ”supervised” refers to the presence of a teacher or supervisor that guides the model's learning process. The process begins with a dataset that consists of input-output pairs, where the inputs are the data features, and the outputs are the corresponding labels or target values. The goal of supervised learning is for the model to learn a mapping or function that can accurately predict or classify new, unseen data based on the patterns and relationships it has learned from the training data. During training, the model adjusts its internal parameters iteratively to minimize the discrepancy between its predictions and the actual labels in the training dataset. Supervised learning is widely applied in various domains, including image recognition[7], natural language processing[8], recommendation systems[9], and many others.

Reinforcement learning is a specific subfield of machine learning that focuses on training agents to make sequences of decisions by interacting with an environment. In reinforcement learning, an agent learns to take actions that maximize a cumulative reward over time. It is different from other machine learning paradigms because it deals with decision-making in dynamic, sequential settings. Some key characteristics of reinforcement learning include exploration vs exploitation, trial-and-error learning, and sequential decision-making. Reinforcement learning has been applied in various domains, including game-playing artificial intelligence (e.g., AlphaGo), robotics[10–12], healthcare[13,14] recommendation systems[15], and autonomous navigation[16,17].

Reinforcement learning operates fundamentally within the framework of a Markov Decision Process, a mathematical formulation that describes an environment in which outcomes are partly random and partly under the control of a decision-maker. A Markov Decision Process is defined by a tuple:

$$(S, A, P, R, \gamma)$$

with S being the set of states, A as the set of actions, P as the transition probabilities, R as the reward function, and γ as the discount factor[18]. In this setup, a reinforcement learning agent learns to make decisions by interacting with its environment, aiming to maximize cumulative reward.

The core of reinforcement learning is to find a policy π, a strategy for selecting actions based on states, that maximizes expected rewards. The value function V^{π}(s) and action-value function Q^{π}(s,a) are crucial, representing the expected cumulative reward from a state or state-action pair, respectively, under a policy π. The agent updates its understanding of the environment and optimizes its strategy using these values, typically employing algorithms like Q-learning, which iteratively adjusts Q values towards optimal rewards. The diagram demonstrating how reinforcement learning works is illustrated in figure 1.

Methodology

To solve the first problem, supervised learning was employed as a dataset containing both text features and labels indicating real or fake news was available. For the second question, as the aim was to approximate human responses and the interaction between the agent and the environment, it was modeled as reinforcement learning.

We aim to determine the extent to which reinforcement learning can be applied. For example, reinforcement learning has been shown to effectively function as a framework for human feedback in previous research[19]. Additionally, large language models (LLMs) like OpenAI's Chat GPT have also demonstrated success[20]. However, we are interested in investigating whether reinforcement learning can still be used as a human

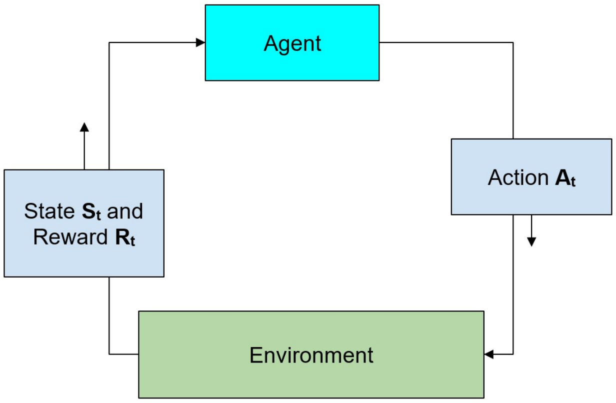

Fig. 1 An illustration showing how reinforcement learning works. The agent passes the action At to the environment, and the environment then passes the state S_t and the reward R_t to the agent, allowing the agent to make better decisions that help it maximize its reward.

feedback framework without requiring the input of the information state (i.e., the text/sentence itself), which is commonly seen in the current research community. This scenario is realistic, as sometimes individuals rely on the reactions of others, including the comments, to determine if a post is real or not. We wanted to explore whether an individual can identify the authenticity of a specific text even without understanding the main posted text. In other words, we wanted to see if we could achieve satisfactory results in modeling/approximating human response, which would ultimately lead to accurate predictions.

Materials and Methods

In this study, we used both supervised and reinforcement learning techniques. The dataset, acquired from Emine Bozkuş and consisting of both real and fake textual information, was originally contained in two CSV files and was transformed into 20-dimensional vectors with values normalized to a range between 0 and 1. Each entry in these vectors represented a ground truth label, where 0 signified fake news, and 1 indicated truth. The dataset was partitioned into training (80%) and testing (20%) sets to facilitate model training and evaluation.

The supervised learning phase involved training a four-layer model using the training dataset. The model consists of three layers, including an input layer with 128 neurons, a hidden layer with 64 neurons—both employing the ReLU activation function—and an output layer with a single neuron using the sigmoid activation function for binary classification. A dropout layer with a rate of 0.3 follows the input layer to mitigate overfitting. The model is compiled with the Adam optimizer, using binary cross-entropy as the loss function and accuracy as the performance metric. It is trained over 10 epochs with a batch size of 64, and employs a 20% validation split. Our goal was to achieve high accuracy in classifying news articles as either real or fake. The trained model was then evaluated using the testing dataset to assess its performance.

For reinforcement learning, the detection of fake news is a challenging problem due to the subjective nature of determining the truthfulness of news. Humans, with their ability to remove biases and interpret text, play a crucial role in making this judgment. Unlike humans, models lack a reference point to verify the authenticity of a piece of information. While they can compare it with existing news sources, there is no guarantee of their authenticity since they are also created by humans. However, humans, as a collective, can identify fake news relatively easily. Therefore, our reinforcement learning model is trained based on human reactions to both fake and truthful news. For simplicity and due to limitations in our research, we use a modified version of rational choice theory and assume that humans are politically neutral and perfectly rational[21].

Proximal Policy Optimization (PPO) is a popular reinforcement learning algorithm used for training deep neural networks to make decisions in various environments, from video games to robotics. The core of PPO is represented by its objective function, which is designed to improve the policy while keeping the updates within a certain range to ensure stability. The equation for PPO's clipped objective function is shown in figure 2.

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\Big[\min(r_t(\theta)\hat{A}_t, \operatorname{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\Big]$$

Fig. 2 $r_t(\theta)$ is the probability ratio of the current policy to the old policy under action at given state s_t, A_t is the advantage estimate at time t, and ε is a hyperparameter, typically set around 0.1 to 0.2, which defines the clip range to prevent too large policy updates. This equation seeks to limit the policy update step to a safe range defined by ε, balancing exploration and exploitation by encouraging moderate changes to the policy[22].

To formulate a fake news detection problem in a reinforcement learning framework, we build two components: agent and environment. Proximal Policy Optimization (PPO) is a common and practical on-policy reinforcement learning algorithm, which can converge faster (9). Therefore, we employed a PPO agent, recognized for its stability and efficiency. We also developed our customized environment upon the OpenAI Gym environment (14), named fake detect gym environment. Specifically, the environment consists of 3 main components: the state, action, and reward function. The state is a 5-dimensional vector initialized with zeros. These dimensions represent different forms of user interactions with the news content: no action, likes, positive comments, negative comments, and retweets. The state vector is a simplified representation of how users engage with a piece of news on a social platform. The observation space is continuous, defined by a Box space that allows each feature to range from 0 to 1. This design choice enables the model to interpret the state as varying levels of engagement intensity or probability.

The action space is discrete, consisting of three possible actions. These actions are quantified as adjustments of -0.1, 0, and +0.1, which correspond to decreasing, maintaining, or increasing the model's confidence in the authenticity of the news content. The actions simulate potential human reactions to the content—like adjusting the belief in its truthfulness based on observed engagement. For example, a significant number of positive comments or likes might increase the model's confidence in the content's veracity, reflected by choosing the +0.1 adjustment. Finally, the reward function aims to quantify the accuracy of the model's predictions. It is calculated based on the difference between the model's confidence in the news being real (as adjusted by the actions taken) and the actual label of the news (real or fake). The closer the model's confidence level is to the true label, the higher the reward. This setup incentivizes the model to accurately predict the authenticity of news articles by aligning its confidence levels closely with the actual truth value of the content. The reward is designed to encourage the model to make adjustments that lead to a more accurate representation of the news content's veracity, using simulated user engagement as a proxy for truthfulness. Using this kind of simulation environment is common in solving reinforcement learning tasks in real scenarios, which can either be transferred to real-world execution via sim-to-real techniques or rely on the integration of real datasets.

We train over 40 episodes, each with 20 steps of interaction with the environment. These interactions were governed by predefined probabilities, adjusted to reflect the authenticity of the text. The reinforcement learning environment incorporated a reward mechanism based on the alignment of the agent's confidence levels with the true labels of the news, introducing randomness to simulate real-world human interaction. The model's performance was evaluated based on the average reward and standard deviation across evaluation episodes. The agent was trained to interact with the environment by taking actions that adjusted its confidence levels by amounts based on fixed probabilities on whether a text is true. In this project, the probabilities used are the following:

If the text is true, then the following adjustment probabilities are used:

$$adjustment\_probs = [0.3, 0.7, 0.9, 0.2, 0.5]$$

Otherwise, these adjustment probabilities are used:

$$adjustment\_probs = [0.7, 0.2, 0.3, 0.5, 0.4]$$

The adjustment probabilities correspond to the actions of no action, likes, positive commenting, negative commenting, and retweeting. For example, if the given text is true, then there is a 30% (corresponds to 0.3) chance that the user takes no action and the no_action value is incremented by 1, and there is a 90% (corresponds to 0.9) chance that the user gives a positive comment and the positive comment action is incremented by 1. The probabilities in this array are arbitrarily decided for simplicity and work limitations, and in the real world, one can substitute more realistic values through observation. We were unable to perform optimization or empirical validation due to limitations in resources.

The reinforcement learning environment included a reward mechanism based on how well the agent's confidence aligned with the true labels of the news articles. The reward formula used is the following:

$$1 - |actual\_label - percentage|$$

where the percentage reflects the model's confidence in the text being true. Additionally, we used a fixed set of probabilities to simulate human interaction with the content, introducing randomness into the environment. The reinforcement learning agent's performance was evaluated based on the average reward and standard deviation across a set of evaluation episodes. The reinforcement learning model was trained and tested on a custom environment, which provided valuable insights into its ability to assess the truthfulness of textual content.

The results from both the supervised and reinforcement learning models will be compared to determine their respective efficacies in solving this challenge.

Results

For the supervised learning model, we were able to achieve an accuracy of over 98% on both the training set and the validation set. Through epochs 1-10, as shown in Figure 3, the training set's accuracy increased until it was over 99.5%. This might indicate that the neural network was overfitting, but the validation set's accuracy hovers around 99%, indicating that no overfitting has occurred. More information on the results is shown in Table 1.

Metric	Training Set	Validation Set
Accuracy	99.33%	96.66%
Precision	99.34%	96.65%
Recall	99.32%	96.60%
F1 Score	99.33%	96.62%

Table 1 The details of the results for the supervised learning model.

Number of episodes trained:	2,000	4,000	2,000,000
Average Reward:	10.45999999046325	10.71000000238418	10.79999998807907
Standard Deviation:	0.728285722004738	0.4711687555274679	0.679705887819817

Table 2 The details of the results for the supervised learning model.

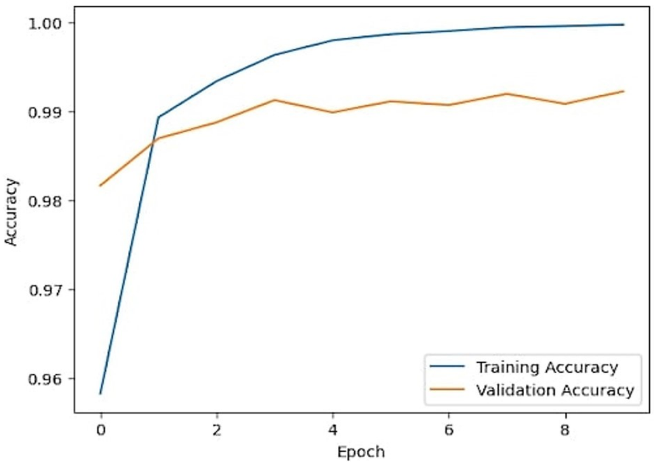

Fig. 3 The results for the supervised learning model with the full dataset. The blue line displays the training accuracy over epochs, while the orange line displays the validation accuracy over epochs.

Conversely, the performance of the reinforcement learning model was underwhelming. The details of the comparison among different training episodes is reported in table 2. No figure is available to visualize the data. After undergoing 2000 iterations of training, it achieved an average reward of 10.46 with a standard deviation of 0.73. To put this in perspective, the model's theoretical maximum reward is 20.0, derived by multiplying 20 (the number of samples per episode) with the maximum attainable reward for each sample, which is 1. Even after an extensive training period of 2 million iterations, the improvement was marginal; the average reward only slightly increased to 10.8 with a standard deviation of 0.68. The model's modest improvement over an extensive training period suggests potential areas for enhancement. These might include adjusting the reward function, refining the state or action space, or exploring more sophisticated learning algorithms to better capture the nuances of fake news detection.

Discussion

The supervised learning model outperformed the reinforcement learning model in terms of consistency and accuracy. The supervised learning model achieved an accuracy of approximately 99%, which is almost perfect. In contrast, even with extensive training, the reinforcement learning model only achieved an average reward of 10.8 out of a theoretical maximum of 20. This indicates that the reinforcement learning model is only slightly better than random chance at correctly identifying fake news or real news, highlighting the need to consider the main text as part of the input.

Some limitations were encountered during the study, including a lack of sufficient resources to train for more than a few million iterations and a scarcity of available datasets for training purposes.

In this work, we have demonstrated the strong performance and generalizability of our supervised learning model. Additionally, we have explored the potential of using reinforcement learning to predict fake news by estimating human responses without incorporating main texts, which, to the best of our knowledge, has not been done before. However, our findings indicate that bypassing the main text as an input for reinforcement learning policy is not sufficient, highlighting the importance of complete data. We suggest that future researchers continue to investigate this problem.

Here are some potential future directions for research:

1. Explore alternative approaches in addition to reinforcement learning.

2. Use different datasets to improve the detection of fake news.

3. Still, use reinforcement learning but with access to limited or different portions of main texts. For example, approximate people's interpretation by considering that some fully understand the main text, some understand only a piece of it, and some do not understand anything.

4. Still use reinforcement learning, but introduce noise to the main text to represent different levels of ease for individuals to understand it.

These extensions go beyond the scope of the current work and can be considered as potential future directions.

References

1 N. Reimann, *Some Americans are tragically still drinking bleach as a coronavirus 'cure'*, www.forbes.com/sites/nicholasreimann/2020/08/24/some-americans-are-tragically-still-drinking-bleach-as-a-coronavirus-cure/, Retrieved from.

2 Z. Khanam, B. Alwasel, H. Sirafi and M. Rashid, *Article 012040*, **1099**, year.

3 J. C. Reis, A. Correia, F. Murai, A. Veloso and F. Benevenuto, *IEEE Intelligent Systems*.

4 A. Mosallanezhad, Proceedings of the ACM Web Conference 2022.

5 Y. Wang, *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, year.

6 E. Bozkuş, *Fake news detection datasets*, www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets, Retrieved from.

7 M. Tan and Q. V. Le, Proceedings of the 36 th International Conference on Machine Learning (ICML.

8 A. Radford and J. W. K. and, Proceedings of the 38 th International Conference on Machine Learning (ICML.

9 J. Yu, H. Yin, X. Xia, T. Chen, J. Li and Z. Huang, *IEEE Transactions on Knowledge and Data Engineering (TKDE*.

10 K. Kumar, I. Essa and S. Ha, *IEEE Robotics and Automation Letters*.

11 H.-L. Hsu, Q. Huang and S. Ha, IEEE International Conference on Robotics and Automation (ICRA.

12 J. Kim and W. Y. and, Conference on Robot Learning (CoRL.

13 Q. Gao, M. Naumann, I. Jovanov, V. Lesi, K. Kamaravelu, W. M. Grill and M. Pajic, ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS.

14 H.-L. Hsu and M. Pajic, *Learning for Dynamics and Control (L4DC*.

15 M. Afsar, T. Crump and B. Far, *ACM Computing Surveys 2022*.

16 Z. Xu, B. Liu, X. Xiao, A. Nair and P. Stone, IEEE International Conference on Robotics and Automation (ICRA.

17 H.-L. Hsu, H. Meng, S. Luo, J. Dong, V. Tarokh and M. Pajic, IEEE International Conference on Robotics and Automation (ICRA.

18 R. Sutton, *Reinforcement learning: An introduction*, MIT Press.

19 L. Kaelbling, M. Littman and A. Moore, *Journal of Artificial Intelligence Research*, **4**, 237–285.

20 J. Achiam, *GPT-4 technical report*.

21 A. Ganti, *Investopedia, Investopedia*, p. – –.

22 T. Simonini, *Hugging Face – The AI Community Building the Future., Hugging Face*, p. – –.