Leveraging RNNs for Detecting Fake News: A Comprehensive Study

Haridas P
School of Computer Science &
Engineering,
Lovely Professional University,
Phagwara, Punjab, IN
hp3.16.bsr@gmail.com

Abstract: In the digital age, the rampant spread of fake news poses a significant threat to the credibility of news sources and the overall integrity of information dissemination. To combat this challenge, our research paper presents an innovative solution that leverages the capabilities of deep learning, specifically using Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) units, in conjunction with GloVe word embeddings. These advanced technologies form the core components of an exceptionally effective system for detecting fake news. Our research findings demonstrate substantial improvements in performance when compared to traditional machine learning models. Notably, our RNN-LSTM-based approach achieves remarkable accuracy and precision, even when confronted with a diverse array of news sources and writing styles

Keyword: - Recurrent Neural Networks, Long Short-Term Memory, GloVe, Deep Learning

INTRODUCTION

Fake news, in its essence, encompasses the deliberate creation of deceptive or falsified information designed to mislead and manipulate individuals, thereby casting detrimental repercussions upon both the individual and the broader society. It exploits the vulnerability of people through the dissemination of fabricated narratives and biased accounts, often for self-serving purposes, and consequently exerts a profound influence on public opinion and societal stability. A stark illustration of this is observed during the COVID-19 pandemic, where the deluge of information, a blend of truth and falsehood, has reached such an extent that the World Health Organization aptly characterized it as an 'information epidemic. The core challenge in fake news detection lies in the extraction and analysis of relevant features within the vast sea of textual data. Traditional methods, reliant on handcrafted feature engineering, often struggle to capture the subtle linguistic nuances and contextual cues indicative of misinformation. Moreover, these traditional Text Classification Techniques [1] while effective to some extent, were notably time-consuming and resource-intensive. The laborious task of manually crafting features for detecting fake news not only incurred substantial time and effort but also incurred significant costs. In contrast, RNNs, specifically designed for sequential data analysis, offer an alternative path. Their ability to model temporal dependencies and context makes them a compelling choice for feature extraction in the realm of fake news detection. RNNs have the inherent capacity to adapt to the complexities of linguistic patterns and evolving narratives present in fake news, alleviating the burdensome task of manual feature engineering. Deep RNN architectures, building upon the foundation of RNNs, extend the capacity of these networks to encapsulate intricate patterns within text data. The hierarchy of layers within deep RNNs enables the automatic discovery of salient features and relationships, thereby enhancing the accuracy and robustness of fake news detection systems. In this context, the combination of RNNs and deep RNNs heralds a promising avenue for feature extraction in the pursuit of more accurate and efficient detection of fake news, while also mitigating the time and resource constraints associated with traditional text classification techniques.

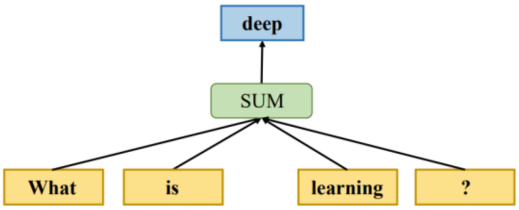
LITERATURE REVIEW

Previous surveys primarily categorized research based on feature perspectives. E.g. Zhou and Zafarani (2018) [2] categorized the approaches for detecting fake news into the four categories listed below: external knowledge-based detection methods, style-based detection methods, propagation-based detection methods, and credibility-based detection methods.

While these categorizations have been valuable, a need has emerged for exploring weakly supervised and unsupervised methods, which have gained prominence due to limited labeled data in real-world applications. These limitations arise from cost considerations and privacy concerns, rendering large-scale labeled datasets elusive. Additionally, it has become apparent that the emphasis on different feature information varies depending on the learning methodology employed—be it supervised, weakly supervised, or unsupervised.

TRADITIONAL APPROACH

Traditional models have advanced text classification, increasing their applicability and accuracy. The initial step involves pre-processing raw input text for training these models, which typically includes tasks like word segmentation, data cleaning, and statistical analysis. Subsequently, text representation techniques are employed to transform the pre-processed text into a computer-friendly format that minimizes information loss. Methods such as Bag-Of-Words (BOW), N-gram, and Term Frequency-Inverse Document Frequency (TF-IDF)[3] are commonly used for this purpose.



(a) CBOW.

Additionally, in our pursuit of improving traditional models for text classification, we incorporated Probabilistic Graphical Models (PGMs) like Bayesian networks to represent the conditional dependencies among various features in graphical form. These models encompass techniques such as Naive Bayes classification.

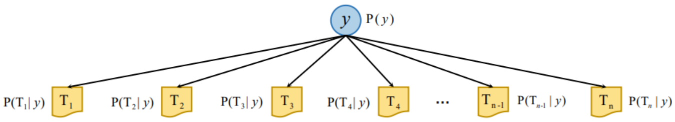
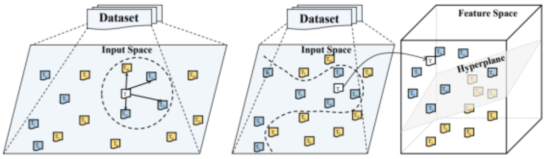


Fig. 3. The structure of Naïve Bayes.

Furthermore, we explored K-Nearest Neighbors (KNN)-based methods as part of our approach. The essence of the KNN algorithm is to classify unlabeled samples by identifying the category with the highest representation among the nearest 'k' samples. KNN offers a straightforward classification approach without the need to construct a formal model and excels in reducing complexity by swiftly determining the 'k' nearest neighbors.

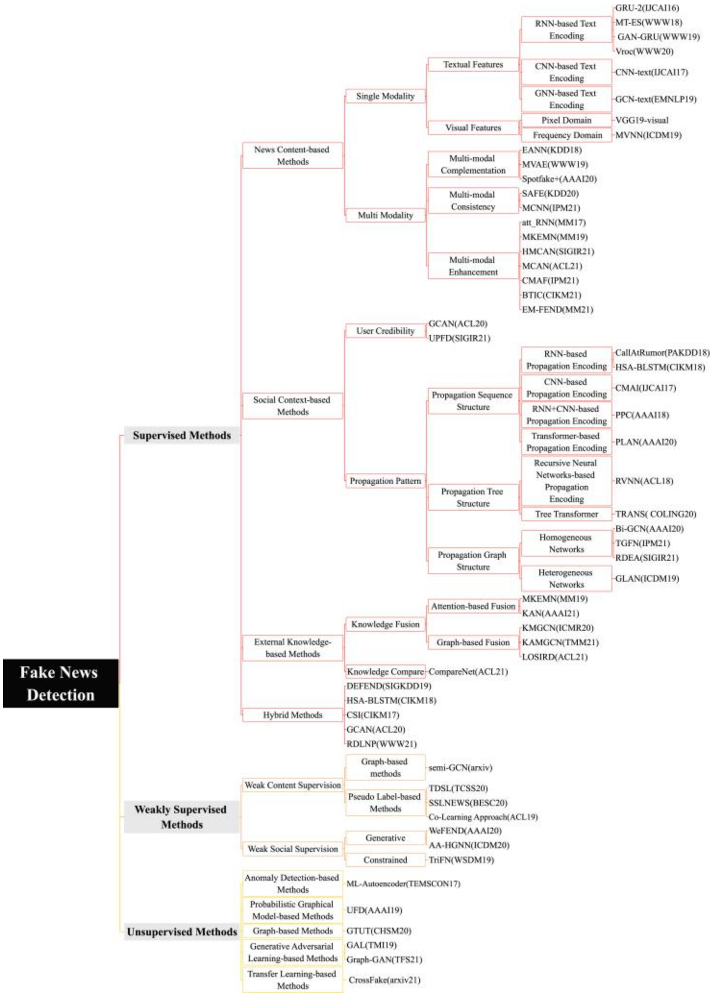


# Support Vector Machines

Moreover, our research encompasses the application of Support Vector Machine (SVM) and Decision Tree-based models to address the challenges of fake news detection. These diverse methods contribute to a comprehensive exploration of traditional and contemporary techniques aimed at enhancing text classification accuracy.

DEEP NEURAL NETWORK BASED APPROACH

Artificial neural networks (ANNs) comprise DNNs, which mimic the functioning of the human brain to automatically extract high-level characteristics from input. This allows the models to outperform traditional models in text reading, image processing, and speech recognition. Data should be classified by analyzing input datasets like an unsupervised, multi-label, single-label, or unbalanced dataset. Based on the characteristic of the dataset, the DNN receives the input word vectors and trains it until the termination condition is attained. The downstream task validates the training model's performance, including Event prediction, question answering, and sentiment classification



In prior research, Dong et al. (2018) applied an attention-based Bi-GRU for news content, a deep neural network for user data, and an attention mechanism to merge text and user information, enhancing fake news detection. Lu and Li (2020) created a fully connected user graph and utilized Graph Neural Network (GNN) [4] techniques to detect fake news through user modeling. Emotions are also pivotal in fake news detection. Zhang et al. (2021) introduced BERT-EMO[5], which considers publisher and social emotions, including their dual emotional interaction, to uncover distinctive emotional cues for fake news detection.

METHODS AND PREPARATION

This research uses "Fake and Real News Dataset" meticulously compiled by Clement Bisallion. This dataset comprises two distinct .csv files, one for real news and another for fake news, housing 21,000 and 23,000 entries, respectively. This extensive dataset offers a comprehensive and diverse collection of news articles that serve as the basis for our investigation.

The data within the dataset is drawn from a variety of American news sources, reflecting the breadth of the American media landscape. The dataset encompasses news articles with content spanning a wide range of themes, from political discourse to everyday news.

Our research methodology was implemented using the Python programming language, which is widely recognized and frequently employed in the realms of data science and machine learning. Throughout the course of the project, PyTorch, a robust and versatile library, played a central and indispensable role. It served to facilitate a spectrum of tasks, encompassing data pre-processing and model training.

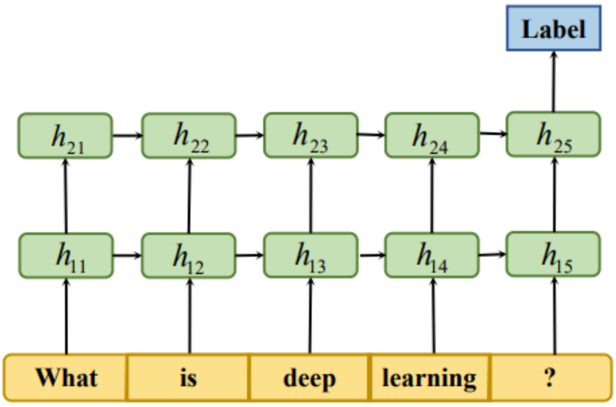
The crucial stage of data pre-processing was initiated to enhance the overall quality and suitability of the dataset. This critical process involved the elimination of irregular characters, data cleaning, and the normalization of textual content. For the purpose of segmenting the text, we adopted the SpaCy Tokenizer, renowned for its reliability and proficiency in text segmentation.

We harnessed the capabilities of the TorchText library, an extension of PyTorch, purposefully designed to manage and manipulate textual data. TorchText greatly expedited the creation of structured, tabular datasets, offering an efficient framework for the organization and management of data. This structured data format significantly facilitated the transition from raw text data to meticulously organized datasets, a pivotal step in the realm of deep learning projects.

In the process of generating data loaders for model training, we implemented Bucket Iterators, an optimization technique adept at batching sequences of varying lengths efficiently. This approach ensured that textual data could be processed effectively during the training phase. To enhance the representation of textual data, we judiciously leveraged GloVe Embeddings, specifically the "glove.6B.100d" embedding matrix. The application of these pre-trained word embeddings served to enrich the model's comprehension of textual content.

At the core of our methodology resides a Bi-directional Recurrent Neural Network with Long Short-Term Memory (Bi-RNN LSTM) architecture. This model is composed of three interconnected layers that span from the input to the output layers. Its bi-directional nature allows it to capture contextual information both in the forward and backward directions, thereby enhancing its ability to understand sequential data.

Recurrent Neural Networks (RNN)



The Recurrent Neural Network (RNN)[6] serves as a prominent tool for capturing long-range dependencies by means of recurrent computation. In the context of text classification tasks, the RNN language model excels in learning historical information, effectively considering the spatial relationships among all words. Initially, each input word undergoes representation as a distinct vector through the utilization of word embeddings. Subsequently, these embedded word vectors are systematically input into RNN cells, sequentially. The output of the RNN cells maintains the same dimension as the input vector and is subsequently directed into the subsequent hidden layer. It is noteworthy that the RNN architecture shares parameters uniformly across various segments of the model, ensuring that the weights assigned to each input word remain consistent. Ultimately, the label assigned to the input text is predicted through the analysis of the final output emerging from the hidden layer.

The training of our model was conducted over three epochs, a widely accepted practice to ensure effective learning from the dataset. For the optimization of deep neural networks, the Adam optimizer was employed, recognized for its efficiency. To address the binary classification nature of our task, we selected the Binary Cross-Entropy with Logits Loss (BCEWithLogitLoss) as the most suitable loss function. This choice aligns with the requisites of binary classification tasks, such as the detection of fake news.

Pseudo-code for the preparation and training:

Step 1: Data Pre-processing

Step 2: Dataset Generation

Step 3: Data Loader Generation

Step 4: Build Text Vocabulary and GloVe Embeddings

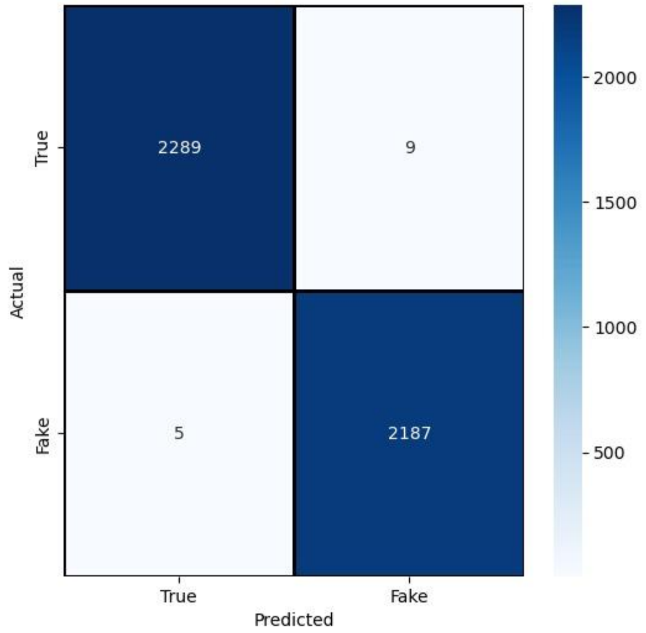
Step 5: Model Architecture

Step 6: Training

This comprehensive methodology, spanning data pre-processing, structured dataset creation, model architecture design, and the training process, serves as the fundamental framework underpinning our research into fake news classification through the application of bidirectional RNN LSTM.

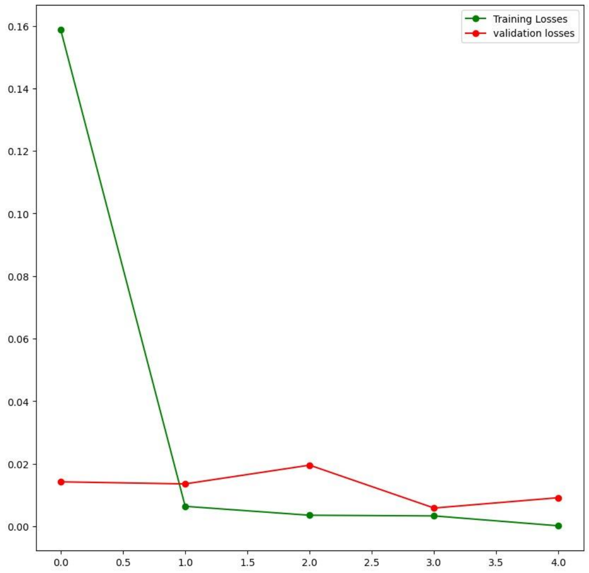
RESULTS AND DISCUSSION

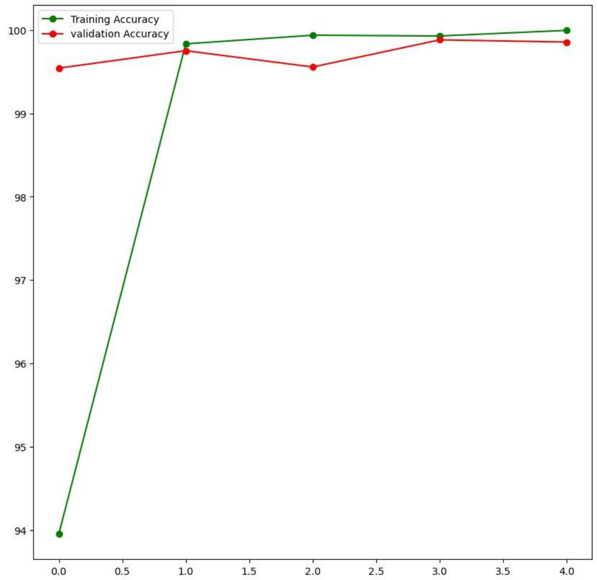
Confusion Matrix:



True Positives denote true news correctly classified model and True negatives denotes fake news correctly classified by the model. While False Positive and False Negatives denote fake news classified as true and True news classified as fake by the model. A total of 14 cases have been misclassified

In case of the False Positives, it is more harmful in terms of Fake news as it deems fake news to be true. The other case is substantially of less relative importance than the discussed one above.





Our Model has been trained with 98.73 % accuracy on Training Data and with 99.88% accuracy on Test Data with no visible case of overfitting.

CONCLUSION

The model yields a result that is both excellent and prone to censorship because of near cent percent accuracy. Since the RNN or any Neural Network is a black box model, explainability of the predictions is something that this model invites openly to. Self or Encoder-Decoder Attention techniques like those in transformer models, model agnostic techniques like LIME and SHAP can be applied to give explanations to the predictions. In future, as a part of the my research, this model would also be fed with generative fake data to check the robustness of the model.

REFERENCES

[1] A Survey on Text Classification: From Traditional to Deep Learning

[2] Deep learning for fake news detection: A comprehensive survey

[3] “Term frequency by inverse document frequency,” in Encyclopaedia of Database Systems

[4] Graph Neural Network for Social Recommendations

[5] Multi-modal Knowledge-aware Event Memory Network for Social Media Rumor Detection

[6] Recurrent Neural Networks