

Comparative Study of Fake News Detection Using Machine Learning and Neural Network Approaches

May Me Me Hlaing¹⁺ and Nang Saing Moon Kham²

University of Computer Studies, Yangon, Myanmar

Abstract. In today's era, fake news related to politic, finance, reputation, health and education gets create and spread like wild fire within the people as the improvement of using social media. Spreading of misinformation is a cause of great concern for human society. Detecting fake news on social media becomes a challenging problem which turns out to be very difficult to manually analyze because more and more online news is increasing on social network. Although a lot of fake news detection researches have shown some significant results and improvements by using different classification algorithms and feature extraction methods, it still has some gaps to meet the important necessities in classifying news. To address this problem, this paper investigates a fake news detection model using machine learning and neural network approaches with frequency-based and word-embedding feature extraction methods. The system performs the experiments upon "ISOT" fake news benchmark dataset. The experimental results show that exploitation of Long Short Term Memory (LSTM) with Glove feature vector achieves better classified results than Feed Forward Neural Network (FFNN), Naïve Bayes (NB) and Support Vector Machine (SVM) in classifying news.

Keywords: misinformation, social media, fake news detection, machine learning, neural network approach

1. Introduction

In modern times, people utilize social media for knowledge, entertainment, education and business purposes. The survey found that globally percentage of internet users in July 2020 stood at 59% of the global population [1]. Fake news aims to mislead the news users' opinion. For a conspicuous example, COVID-19 pandemic quickly spreading around the world in this period and fake information related to this disease gets created in our society. Such misinformation caused anxiety among people and even deadly consequences in health problem. As another instance, in our country, during the 2020 election, massive fake information diffused on social media platform which aimed to mislead opinion among the citizens, organizations and election parties. Online news is noisy, incomplete and unstructured format. Thus, feature representation method plays a vital role in classifying news is fake or not. In this paper, we presented frequency-based method and pre-trained word embedding method for feature extraction.

The main objective of this paper is to analyse comparative study of fake news detection using machine learning and neural network approaches. It also aims to know the ideas and process of fake news detection system and explore which methods are more efficient and effective for news classification. We have divided the system into four main phases: data acquisition, data pre-processing, feature extraction and system modeling. After that, we have evaluated and compared the performance results using different evaluation metrics and analysed which classification model achieves better result than other models.

The remainder of the paper is organized as follows. In section II, we review related work of the system. The explanation of proposed system and methodology are provided in section III. We present characteristics of the ISOT fake news dataset, development setting and experimental results with evaluation metrics in section IV. The paper is concluded in section V by glimpsing our future direction about the current system.

⁺ Corresponding author. Tel.: +95-9-5110568; fax: +95-1-610633
E-mail address : maymimihlaing88@ucsy.edu.mm

2. Related Work

There persists a sizeable body of research for fake detecting system using supervised machine learning techniques and deep learning methods. In this section, some of the research works about this problem are reviewed the following.

The authors applied Naïve Bayes classifier to detect fake news based on BuzzFeed dataset in [2]. It also described the nature and characteristics of fake news and similarity between spam messages and fake news. In reference [3], the authors proposed a classification model to detect fake news using five classifiers with different NLP feature representation methods such as TF-IDF, count vectorizer and word2vec. It also explored which feature representation method is most suitable for contextual information in fake news dataset. The authors described a fake news detection model using six supervised machine learning methods with TF-IDF N-gram analysis based on a news benchmark dataset and compared the system performance based on these methods [4]. In reference [5], the authors proposed a fake news detection model using four different machine learning techniques with two word embedding methods (Glove and BERT) to detect sarcasm in tweets. The authors demonstrated an automated fake news detection system using machine learning techniques based on three distinct feature extraction techniques : TF-IDF using bi-gram frequency, probabilistic context free grammars(PCFGs) and combined feature and the result indicated that PCFGs are good union in classifying news is fake or not in [6].

3. Proposed System and Methodology

This section stands for description of overall proposed system design and detailed explanation of methodology used in this system. The overview flow diagram is first described in Fig 1. After that, the explanation of methodology in each step as follows.

3.1. Data Preprocessing

Data pre-processing phase is very important step in any data science lifecycle, as pre-processed data will make that the data is reliable and consistent in building machine learning models. Social media news is unstructured, big and noisy. Thus, these data need to be pre-processed and the briefly description of data pre-processing steps is described in the following.

In data pre-processing step, we have first dropped duplicate rows and missing values from the dataset. After that, we have removed punctuations, numbers, alphanumeric texts, non-English words and stop words from the news articles. These values can cause the model over fitting the data and these are not added any real values to the actual news content. Moreover, we have performed tokenization process to split the stream of sentence into words phrases, symbols or other meaningful information. Then, we have also applied lemmatization technique to transform the tokenize words into standard form and this technique reduces inflectional forms and changes derivationally relates forms of the words to the common base form.

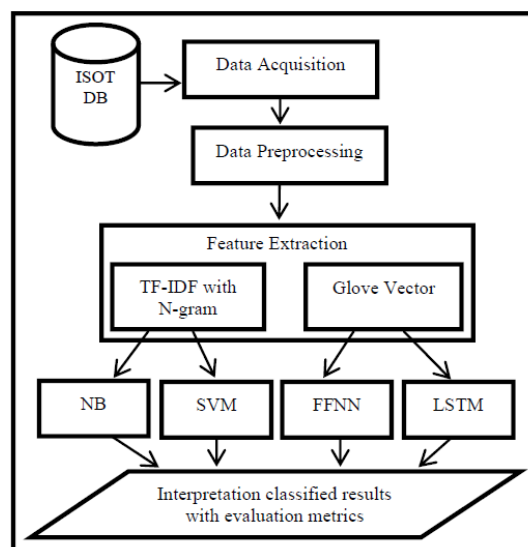


Fig. 1: Overall Design of the Proposed System

3.2. Feature Extraction

Machine learning model can deal with numeric values only. Therefore, we need to transform the text input data into numeric vector form for applying classification methods. The process of converting text data into numerical representation is called vectorization. There are many word vectorization techniques in text processing. Among them, this paper presents frequency-based method (TF-IDF) and pre-trained word embedding method (Glove). The explanation of these methods are described the following.

TF-IDF with N-gram Analysis: TF-IDF is sparse vector representation and it evaluates how relevant a word in a collection of documents or corpus [7]. TF-IDF techniques use all the tokens in the sentence as vocabulary. TF means term frequency which can use to measure how frequently a term in one document. IDF measures the important of a term in document. The value of TF-IDF weight can be obtained as a way of normalization in Equation (1), where TF is term frequency, d is a document, IDF is inverse document frequency and D is corpus or documents in the dataset.

$$TF - IDF(t, d, D) = TF(t, d).IDF(t, D) \quad (1)$$

N-gram is a continuous sequence of items with length n. Using N-gram causes the classifiers more restrictive. In this system, we utilized TF-IDF feature extraction method with n-gram analysis and analysed how effect on the system performance by varying n-gram length and TF-IDF feature size.

However, TF-IDF method only considers directly in the word-count space and it cannot capture position in sentence, word meaning, synonym and co-occurrence in different documents. To tackle these problems, the researchers have been used pre-trained word embedding methods to generate more efficient feature vector. In the proposed system, we also described Glove word embedding method to convert feature vector.

Glove: It stands for global vector for word representation and it is based on matrix factorization method. Glove vector can capture the semantics of analogy. Training is performed on the aggregate global word to word co-occurrence statistics from a corpus, and the resulting word representation vector is linear structure. In Glove dataset contains pre-trained English word vectors on the combined Wikipedia. There are four word vector representation files in Glove: glove 6B.50d.txt (6 billion token and 50 features), 6B.100d.txt (6 billion token and 100 features), 6B.200d.txt (6 billion token and 200 features) and 6B.300d.txt (6 billion token and 300 features). In this system, we used Glove 6B.50d.txt, 6B.100d.txt and 6B.200d.txt for feature representation. Glove model was released in 2014 by Stanford University [8].

3.3. Modeling

Some of the most popular supervised learning algorithms for text classification model include SVM, the family of Naïve Bayes algorithms and deep learning methods [9]. We studied in this research four different classification algorithms namely, NB, SVM, FFNN and LSTM. These methods are briefly described as follows.

NB: It is a simple probabilistic classifier based on applying Bayes theorem with strong independence assumptions between features. To classify the text data, it calculates their co-occurrences and finds it their probability compared with its class labels as Equation (2), where x_i means the number of words found in a text (news in ISOT datasets) and c is binary class labels (real or fake). According to Equation (2), if probability value of the class label ($P(c)$), which targets to “Real” is higher than that of “Fake”, Naïve Bayesian concludes that the input news is “Real”. Otherwise, it was regarded as “Fake”.

$$P(c | x_i) = \frac{P(x_i | c)P(c)}{P(x_i)} \quad (2)$$

SVM: Regarding SVM classification algorithm, we use binary classification that separates the datasets attributes linearly by finding the hyper plane calculated according to Equation (3), where w are the weight values for input words (x), w_0 is the initialized weight and w^T are the transposed vector that are changing the value depending on the value of x . Function $f(x)$ is the boundary function using the parameters. In our system, we use binary SVM algorithm to discriminate the input data text that they belong to “Real” or “Fake” news according to the training data from the dataset.

$$f(x) = w^T x_i + w_0 \quad (3)$$

Neural Network: Unlike the machine learning algorithm, the structure of neural network is quite complicated and it is composed of several layers of interconnected nodes. There are different kinds of neural networks such as feed forward neural network (FFNN), radial basic function neural network, multi-layer perception, convolutional neural network, recurrent neural network – long short term memory (LSTM) and modular neural network [10]. Among them, we applied FFNN and LSTM to classify news articles in the proposed system.

FFNN: It is an artificial neural network and use for classification problems on time series and sequential data. In its structure, there is no feedback action between layers and it includes an input layer, an output layer and one or more hidden layer [11]. In this system, we used two-layer FFNN with sigmoid activation function.

LSTM: It is a special type of recurrent neural network that process the sequential data and time series data. It is more suitable for text classification rather than machine learning algorithms because LSTM can capture long-term dependencies between word sequences and it composed of three gates. A cell state helps the data to flow through the cell units without being changed by allowing only a few linear interactions. The input gate in LSTM controls the new values flow into the cells, the forget gate control the extent to which a value remains in the cell and the output gate use to compute the activation output of the LSTM cell [12]. In the proposed model, we applied sigmoid activation function and Adam optimization function to classify news.

4. System Evaluation

In this section, we first presented the characteristics of ISOT fake news dataset. Next, system development setting is described in sub-section. Then, we compared the system performance results based on two different feature representation techniques with machine learning and neural network approaches.

4.1. ISOT Dataset

It contains two types of news articles: Real and Fake during 2016 and 2017. This dataset is launched from the Information Security and Object Technology (ISOT) Research Lab. In this dataset, real news articles are collected from “Reuter” news website. As for the fake news articles, they were collected from unreliable websites which were identified by Wikipedia and PolitiFact fact checking website [13]. The detailed description of news categories and numbers of articles are described in Table 1.

Table 1: News Categories and Number of News Articles per Category

News	No.of articles	Subjects	
Real News (2016-2017)	21417	World News	10145
		Politic News	11272
Fake News (2016-2017)	23481	World News	16640
		Politic News	6841

4.2. Development Setting

The proposed system is implemented with python programming. Numpy and Pandas libraries are used to manage data preparation and file handling. Scikit-learn, Keras and Tensorflow libraries are exploited to apply NB, SVM, FFNN and LSTM. Due to the limitation of computing resource, we only use politic news articles from dataset in this system. We split the data into 20% for testing and 80% for training. The parameters for four classification algorithms, we use default parameters included in Keras and Tensorflow libraries because these parameters setting are simplest and powerful for classification purpose. The performance measure is reported by 5-fold cross validation in this system.

4.3. Experimental Results

In system implementation, we first demonstrated using two machine learning models: NB and SVM using TF-IDF method with various N-gram lengths from n=1 to n=4. Furthermore, we analysed the results varying the number of feature vector size in TF-IDF method. The obtained results are depicted in Fig. 2.

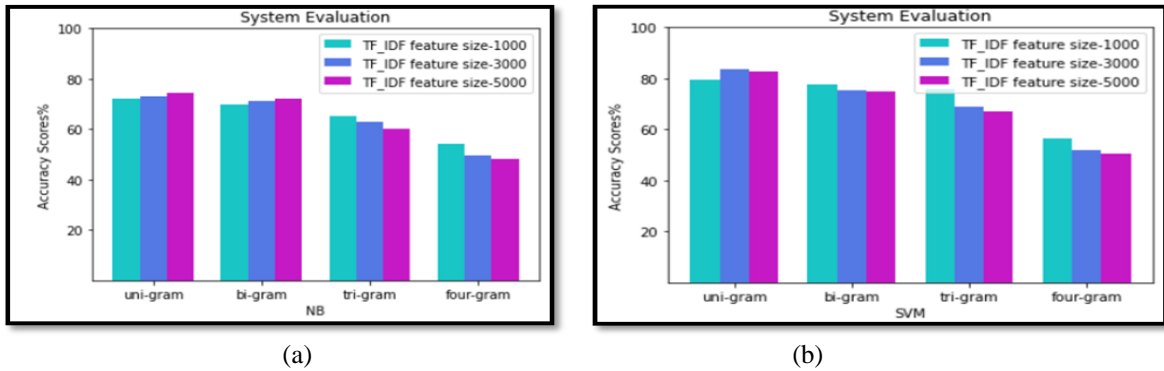


Fig. 2: (a) Accuracy of the system using NB and (b) Accuracy of the system using SVM

From the results, we found that SVM classifier achieved better result than NB in all different feature sizes. The highest accuracy 83% was obtained using SVM with uni-gram and 3000 TF-IDF feature vector. The lowest accuracy 48% was achieved using NB with four-gram and 5000 TF-IDF feature vector. We can consider SVM classifier is more suitable for longer text classification (news) rather than NB and both classifiers perform well on matter the number of TF-IDF feature sizes used. However, the accuracies of NB and SVM slightly decrease when increasing n-gram length especially in tri-gram and four-gram.

In neural network approach, we investigate the classification accuracy using FFNN and LSTM approaches with Glove feature vector. In our experiment, we used three word vector representation files in Glove (6B.50d.txt, 6B.100d.txt and 6B.200d.txt). The exploration results are showed in Fig. 3.

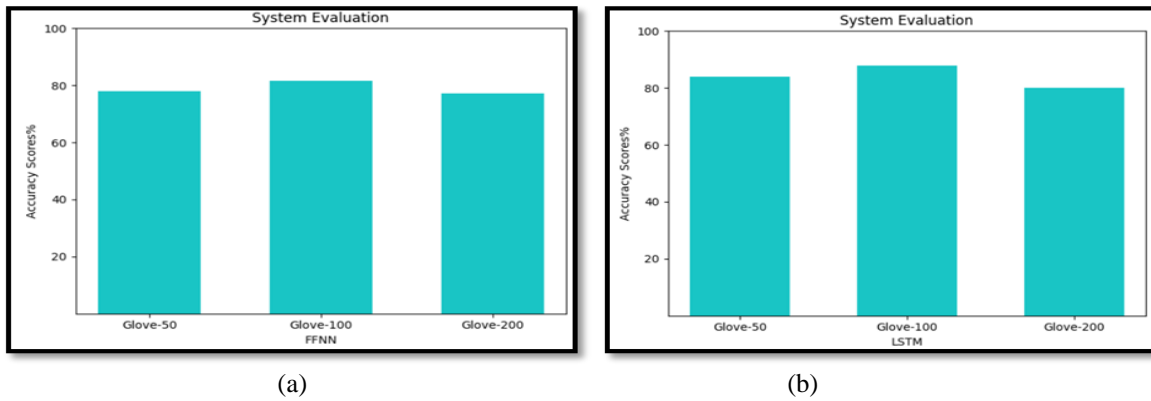


Fig. 3: (a) Accuracy of the system using FFNN and (b) Accuracy of the system using LSTM

From the obtained results in our experiment, LSTM model achieves better accuracy than FFNN in all Glove feature vectors and both classifiers perform well on all Glove features vector sizes. Moreover, LSTM model gets efficient accuracy than NB and SVM.

4.4. Evaluation Metric

Classification process can cause imbalanced classification problem when the class distributions are not equal in dataset. Accuracy measure of classification results alone can be misleading due to this situation. Thus, we also evaluate the system performance using precision, recall and F1 score based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values. These values are derived from confusion matrix from each algorithm and these matrixes show right or wrong classified results such as TP, TN, FP and FN. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. Recall refers to the percentage of the total relevant results correctly classified by algorithm. F1 score is the weighted average of precision and recall. The overall performance is depicted in Fig.4 with precision, recall and f1 measure. In the figure, LSTM with Glove feature vector could stand tallest around 86% and 88% in precision and recall. Based on the result, we can also conclude Glove vector with LSTM model performs better than other algorithms in all evaluation metrics because Glove word representation vector capturing the semantic analogy rather than TF-IDF method and LSTM perform well in sequence data and it can consider the previous data stored memory rather than NB, SVM and FFNN.

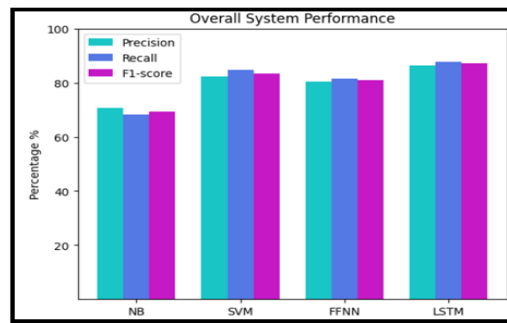


Fig. 4: Overall Performance of the System

5. Conclusion

Due to development of the latest technology, online users are increasing day by day in today's world. At the same time, fake news is widely spreading among online news consumers to mislead their opinion on politics, finance, education and healthcare. Fake news can cause serious harmful impact on human community. Thus, classifying news on social media is an emerging problem in today's research area. In this paper, we presented news classification system using different feature representation techniques with machine learning and neural network approaches. The experimental results have shown that LSTM with Glove vector model outperforms other classification models. The system only applied one offline benchmark dataset "ISOT" for the experiment. As a future work, we plan to implement real time fake news classification system using News API and Twitter API. We also aim to extend the proposed system using transformers such as "BERT", "RoBERT" and "OpenAI's GPT-3" to analyse the news article is real or fake.

6. References

- [1] <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- [2] Granik, M., & Mesyura, V. (2017, May). Fake news detection using naive Bayes classifier. In 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) (pp. 900-903). IEEE.
- [3] Vijayaraghavan, S., Wang, Y., Guo, Z., Voong, J., Xu, W., Nasser, A., ... & Wadhwa, E. (2020). Fake News Detection with Different Models. arXiv preprint arXiv:2003.04978.
- [4] Ahmed, H., Traore, I., & Saad, S. (2017, October). Detection of online fake news using n-gram analysis and machine learning techniques. In International conference on intelligent, secure, and dependable systems in distributed and cloud environments (pp. 127-138). Springer, Cham.
- [5] Khatri, A. (2020). Sarcasm detection in tweets with BERT and GloVe embeddings. arXiv preprint arXiv:2006.11512.
- [6] Gilda, S. (2017, December). Evaluating machine learning algorithms for fake news detection. In 2017 IEEE 15th Student Conference on Research and Development (SCOREd) (pp. 110-115). IEEE.
- [7] Liu, Q., Wang, J., Zhang, D., Yang, Y., & Wang, N. (2018, December). Text features extraction based on TF-IDF associating semantic. In 2018 IEEE 4th International Conference on Computer and Communications (ICCC) (pp. 2338-2343). IEEE.
- [8] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
- [9] Aas, K., & Eikvil, L. (1999). Text categorisation: A survey.
- [10] Wehle, H. D. (2017, July). Machine learning, deep learning, and ai: What's the difference?. In International Conference on Data scientist innovation day, Bruxelles, Belgium.
- [11] Lomuscio, A., & Maganti, L. (2017). An approach to reachability analysis for feed-forward relu neural networks. arXiv preprint arXiv:1706.07351.
- [12] Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.
- [13] https://www.uvic.ca/engineering/ece/isot/assets/docs/ISOT_Fake_News_Dataset_ReadMe.pdf