Multimedia Tools and Applications (2022) 81:19341–19349
https://doi.org/10.1007/s11042-021-11065-x

1182: DEEP PROCESSING OF MULTIMEDIA DATA

A novel self-learning semi-supervised deep learning network to detect fake news on social media

Xin Li[1] · Peixin Lu[1] · Lianting Hu[1] · XiaoGuang Wang[1] · Long Lu[1]

Received: 24 July 2020 / Revised: 21 April 2021 / Accepted: 6 May 2021 /
Published online: 2 June 2021

Abstract
Social media has become a popular means for people to consume and share news. However, it also enables the extensive spread of fake news, that is, news that deliberately provides false information, which has a significant negative impact on society. Especially recently, the false information about the new coronavirus disease 2019 (COVID-19) has spread like a virus around the world. The state of the Internet is forcing the world's tech giants to take unprecedented action to protect the "information health" of the public. Despite many existing fake news datasets, comprehensive and effective algorithms for detecting fake news have become one of the major obstacles. In order to address this issue, we designed a self-learning semi-supervised deep learning network by adding a confidence network layer, which made it possible to automatically return and add correct results to help the neural network to accumulate positive sample cases, thus improving the accuracy of the neural network. Experimental results indicate that our network is more accurate than the existing mainstream machine learning methods and deep learning methods.

Keywords Fake news · Social media · Semi-supervised deep learning network · Confidence values

✉ Long Lu
lulong@whu.edu.cn

Xin Li
XinLi2020@whu.edu.cn

Peixin Lu
Lupx@whu.edu.cn

Lianting Hu
LiantingHu@whu.edu.cn

XiaoGuang Wang
whu_wxg@126.com

[1] School of Information Management, Wuhan University, Wuhan, China

1 Introduction

Social media has become a major way of news consumption mainly because it is free and easy to access, and can rapidly spread posts. Therefore, it is an excellent way for individuals to obtain and publish various kinds of information [13–15]. However, the quality of news works on social media is often lower than that of traditional news sources because the contents on social media cannot be effectively supervised [2, 10, 15]. In other words, social media also allows fake news to extensively spread. Especially recently, the false information about the new coronavirus disease 2019 (COVID-19) has spread like a virus around the world. The state of the Internet is forcing us to take unprecedented actions to protect the "information health" of the public [15, 16].

It is important but challenging to find out wrong information on social media partially because human eyes are not able to distinguish true news from fake ones [11]. To facilitate the study of fake news, researchers have presented many fake news datasets such as *BuzzFeedNews, LIAR* [18], *CREDBANK* [7], *BuzzFace* [12], *FacebookHoax* [17], *and FakeNewsNet* [13–15], which contain the linguistic and social context features of social media content. Despite the existence of multiple fake news datasets, a comprehensive and effective computational solution for detecting fake news has become one of the major obstacles.

Although there are many fake news data sets available, a comprehensive and effective algorithm for detecting fake news has become one of the major obstacles. The existing research on false news detection can be roughly divided into two categories, namely, supervised learning methods based on machine learning [1, 3, 5, 8, 9], and supervised learning methods based on deep learning [4, 6, 19–21]. These models have achieved some results in various false news detection datasets. Shu et al. [15] applied standard machine learning models including support vector machines (SVM), logistic regression (LR), Naive Bayes (NB), and Convolutional Neural Network (CNN), provides similar results around 58% to 63%. Jwa [6] applied the Bidirectional Encoder Representations from Transformers model (BERT) model to detect fake news with a accuracy around 75%. An ensemble learning model combining four different models is proposed for fake news detection, and a higher accuracy of 72.3% is obtained in [4].The accuracy score obtained by FakeDetector with Deep Diffusive Network Model (DDNM) in [21] is 0.63. A model named as TI-CNN (Text and Image information based Convolutinal Neural Network) is proposed in [20]. By projecting the explicit and latent features into a unified feature space, TI-CNN is trained with both the text and image information simultaneously. However, most of methods around fake news detection threats it a supervised learning problem: given an existing dataset of fake news, train a classifier such that it can accurately predict the authenticity of news. In fact, annotated datasets are rare and hard to obtain as fake news circulates through websites. In addition, supervised learning model cannot achieve self-learning as it ignores the correlation between real and false data.

Therefore, learning from some state-of-the-art methods, our work aims to study a self-learning semi-supervised deep learning network that trains supervised and unsupervised tasks simultaneously to detect fake news on social media, and compare the results with existing supervised learning methods. Specifically, the work of this thesis is as follows: **1.** design a semi-supervised deep learning network that simultaneously trains supervised and unsupervised tasks using modified deep learning machines; **2.** make it possible to automatically add highly accurate unlabeled data to the training set and continuously expand the training set in the

multi-iterative training process to achieve self-learning; **3.** Compared with existing machine learning methods and deep learning networks, especially in cases of incomplete annotated training datasets or relatively small datasets, the performance of our method has been improved. In particular, its performance has been improved by 10% compared with that of neutral networks and even more compared with that of machine learning methods.

2 Self-learning semi-supervised deep learning model

Figure 1 shows the workflow of our paper, a) data collection process in this paper, b) semi-supervised self-leaning deep learning model which simultaneously trains supervised and unsupervised tasks using a modified deep learning machine L. The former involves training a supervised learning machine that requires only a small portion of labeled data, while the latter predicts the remaining unlabeled parts and returns a highly confident pseudo label of unlabeled data to enrich labeled datasets.

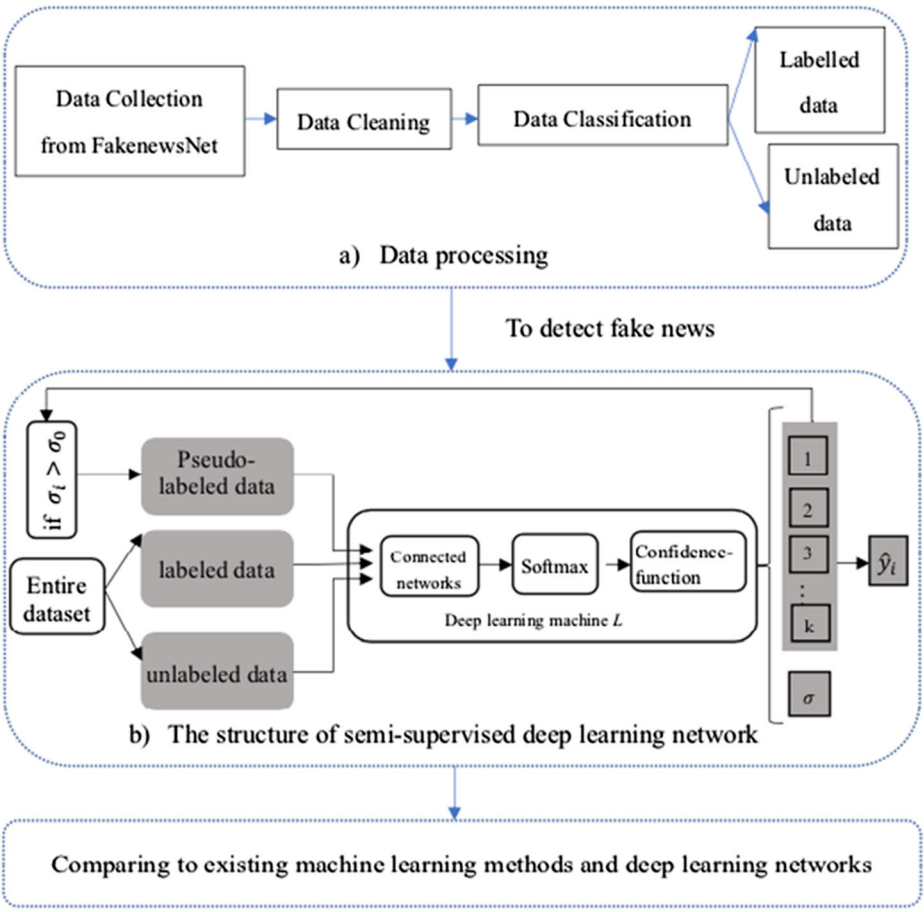

Fig. 1 The workflow of our paper

2.1 Model training process

D_l denotes the labeled examples in trained dataset with a size of $|L|$, $D_l^0 = \{(X1, y1), (X2, y2), \ldots, (Xl, yl)\}$, and D_u denote the unlabeled examples in test dataset with a size of $|U|$, $D_u = \{Xl+1, Xl+2, \ldots, Xl+u\}$. As shown in Fig. 1b), the workflow of the self-learning semi-supervised deep learning machine can be described as follows:

Initialize:

In the supervised deep learning module, D_l^0 is used as a training set to train the deep learning machine L. Then, in the unsupervised deep learning module, the pseudo-labels of $D_u^{'} = \{(X_{l+1}, \widehat{y}_{l+1}), (X_{l+2}, \widehat{y}_{l+2}), \ldots, (X_{l+u}, \widehat{y}_{l+u})\}$ are generated by the trained deep learning machine L and their confidence values σ. If σ_0 is the threshold to filter the unconfident pseudo labels in $D_u^{'}$, then the confident pseudo label set of $D_u^{'}$ can be expressed as $D_{pseu}^0 = ((X_{l+i}, \widehat{y}_{l+i}), (X_{l+i+1}, \widehat{y}_{l+i+1}), \ldots, (X_{l+p+i}, \widehat{y}_{l+p+2}))$ with a size of $|P_0|$.

Repeat:

Then, the new training set $D_l^1 = \left|D_l^0 \cup D_{pseu}^0\right| = \left\{(X_1, y_1), (X_1, y_1), \ldots, (X_l, y_l), \ldots, \left(X_{l+p}, y_{l+p}\right)\right\}$ is used to retrain the deep learning machine L to generate new confident pseudo label set D_{pseu}^2 with a size of $|P_l|$ and a new training set $D_l^2 = \left|D_l^1 \cup D_{pseu}^1\right|$. Repeat this step until $D_{pseu}^t = D_{pseu}^{t+1}$. The experiments proved that this algorithm converges to the optimal solution at a greater speed.

2.2 The basic architecture of deep learning machine *L*

The deep learning machine L is constructed by adding a confident-level layer to existing neural networks, such as recurrent neural networks (RNN), CNN, long short-term memory (LSTM) and BI-LSTM. Here, we take BI-LSTM as an example to introduce the architecture of deep learning machine L. The major components of the deep learning machine L are described below:

2.2.1 Token embedding layer

The token embedding layer maps each token in the input sequence to a token embedding. The word vectors are trained in more ways like one-hot embedding, distributed representation, Neural Network Language Models, word2vec, BERT etc. Word2vec was selected as our token embedding layer in this work. Extracting a text sequence $S = \omega_0, \omega_1, \ldots, \omega_s$ from a collection, if the forward calculation process of the Skip-gram models is written in mathematical form, we get:

$$\mathrm{p}(\omega_0|\omega_i) = \frac{e^{U_0 \cdot V_i}}{\sum_j e^{U_j \cdot V_i}},$$

where, V_i is a column vector of the matrix in embedding layer, also be called the input vector of ω_i. U_j is a row vector of the matrix in softmax layer, also known as the output vector of ω_i.

The loss function of the Skip-gram models is obtained by adding the probability of positive and negative examples in target corpus by using binary logistic regression.

$$J(\theta) = \log \mu(U_0 \cdot V_i) + \sum_{j=1}^{k} E_{\omega_j \sim p_n(\omega)} \left[log\mu\left(-U_j \cdot V_i\right) \right].$$

2.2.2 Dropout layer

Dropout layer is a simple way to prevent neural networks from overfitting. Large networks are also slow to use, and thus difficult to deal with overfitting by combining the predictions of many different large neural nets during testing. The key idea of Dropout is to randomly drop units (together with their connections) from the neural network during training, which prevents units from excessive co-adapting. The random sampling probability was set to 0.5 in this paper, and the sampling probability can also be determined by the verification set.

2.2.3 BI-LSTM layer

In this subsection, we start with a brief review of the fundamentals of BI-LSTM networks, which are a type of RNN and composed of forward LSTM and backward LSTM. LSTM selectively forgets part of the historical information through three gates (input gate, forget gate and output gate), adds part of the current input information, and finally integrates it into the current state to generate the output state. It takes a sequence $\{x_s\}_{s>1}^{S}$ of length S as its input and outputs a S-long sequence of $\{h_s\}_{s>1}^{S}$ hidden state vectors using the following equations:

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ C_t &= f_t * C_{t-1} + i_t * tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t * tan\, h(C_t). \end{aligned}$$

Where $h_0 = 0$. The sigmoid σ and *tanh* functions are applied element-wise. The W matrices and b vectors are the trainable parameters of the LSTM.

2.2.4 Softmax layer

Using a fully-connected neural network, the label prediction layer maps the output from the token BI-LSTM layer to a sequence of vectors containing the probability of each label for each corresponding token. The softmax layer, one of the most popular units for neutral network, is used for multi-classification in this paper. It maps the output of many neurons into the interval of (0,1), which can be understood as the probabilities of multi-classification. For $D_l^0 = \{(X_1 y_1), (X_2, y_2), \ldots, (X_l, y_l)\}$ with k classifications $y^{(i)} \in \{1, 2, 3, \ldots, k\}$. For every input X_i,

$$p(y = j|X_i) = \begin{bmatrix} p\left(y^{(i)} = 1 \middle| X_i; \theta\right) \\ p\left(y^{(i)} = 2 \middle| X_i; \theta\right) \\ \vdots \\ \left(y^{(i)} = k \middle| X_i; \theta\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T \cdot X_i}} \begin{bmatrix} e^{\theta_1^T \cdot X_i} \\ e^{\theta_2^T \cdot X_i} \\ \vdots \\ e^{\theta_k^T \cdot X_i} \end{bmatrix}.$$

2.2.5 Confidence-function layer

As described in Section 2.1, the confidence-function layer is set to calculate the confidence value σ of each element in D_u, and generate the pseudo labels in $D_u^{'}$. For every input X_i,

$$\sigma_{X_i} = max(0, p(y = j|X_i)\Big),$$

Suppose σ_0 is the threshold to filter the unconfident pseudo-labels in $D_u^{'}$, then the confident pseudo-label of an element X_i in $D_u^{'}$ c, if $\sigma_{X_i} > \sigma_0$, then,

$$\widehat{y}_i = \begin{cases} 1, & if\ j = argmax\, p(y = j|X_i\Big) \\ 0, & otherwise \end{cases}.$$

Then we obtain the whole confident pseudo-label set of $D_u^{'}$, D_{pseu}^0 with a size of $|P_0|$. Then, the new train set $D_l^1 = \left|D_l^0 \cup D_{pseu}^0\right| = \{(X_1, y_1),\ (X_1, y_1),\ \ldots,\ (X_l, y_l),\ \ldots,\ (X_{l+p}, y_{l+p})\}$ is used to retrain the deep learning machine L, to generate a new confident pseudo label set D_{pseu}^2 with a size of $|P_l|$ and a new training set $D_l^2 = \left|D_l^1 \cup D_{pseu}^1\right|$. Repeat this step until $D_{pseu}^t = D_{pseu}^{t+1}$.

3 Experiments and results

3.1 Materials and datasets

The fake news data repository FakeNewsNet consists of two comprehensive datasets, each featuring news content, social context, and spatiotemporal information, which were released in 2019 and are also being constantly updated. The latest update version of PolitiFact and GossipCop datasets from FakeNewsNet repository was used to detect fake news in this paper.

3.2 Evaluation metrics

To evaluate the performance of the self-learning semi-supervised deep learning model and compare with other existing machine learning and deep learning methods, we use precision, recall and F1-measure as experiment metrics:

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN},$$

$$F_1 = \frac{2 * precision * recall}{precision + recall}.$$

Here, True Positive (TP) equals the number of data that are correctly identified. False Positive (FP) equals the number of data which are mistakenly identified. False Negative (FN) is the number of data which are not identified.

3.3 Experimental results

We evaluated the performance of the self-learning semi-supervised deep learning model by comparing with machine learning methods such as SVM and NB, and deep learning methods, BI-LSTM network, CNN. Our methods used L as deep learning machine, respectively. We used the default settings provided in the scikit-learn, without tuning parameters.

As shown in Table 1(a), when we used 80% of labeled data for training and 20% of unlabeled data for testing, the self-learning semi-supervised deep learning model based on L achieved a precision of 0.90, a recall score of 0.86, and a F1-score of 0.88, respectively, demonstrating the best performance. The results of the deep learning methods were not significantly different, but were about 30% better than those of the machine learning methods. As shown in Table 1(b), when we used 50% of labeled data for training and 50% of unlabeled data for testing, the precision of our method was 0.88, about 30% higher than that of machine learning methods and 10% higher than that of deep learning methods.

As shown in Table 1(c), when we used 20% of labeled data for training and 80% of unlabeled data for testing, the precision of our method was 0.88, about 40% higher than that of machine learning methods and 15% higher than that of deep learning methods, which proved that our method performed better and more consistently in the case of incomplete annotated training datasets or relatively small datasets than supervised learning methods such as deep learning models and machine learning methods.

4 Conclusion

The fast spread of fake news has raised concerns around the world recently. These fake political news may have severe consequences, the identification of them grows in importance.

Table 1 (a) Experimental results when 80% of labeled data were used for training and 20% of unlabeled data for testing; (b) Experimental results when 50% of labeled data were used for training and 50% of unlabeled data for testing; (c) Experimental results when 20% of labeled data were used for training and 80% of unlabeled data for testing

	Precision	Recall	F1
(a)			
SVM	0.66	0.69	0.61
NB	0.62	0.57	0.61
CNN	0.83	0.77	0.79
BI-LSTM	0.85	0.84	0.86
Our method	0.90	0.86	0.88
(b)			
SVM	0.63	0.59	0.55
NB	0.56	0.53	0.55
CNN	0.75	0.77	0.79
BI-LSTM	0.79	0.81	0.82
Our method	0.89	0.83	0.85
(c)			
SVM	0.56	0.49	0.45
NB	0.58	0.59	0.54
CNN	0.71	0.69	0.79
BI-LSTM	0.73	0.71	0.74
Our method	0.86	0.83	0.81

In this paper, we designed a self-learning semi-supervised deep learning network to detect fake news on social media. A confidence network layer automatically returns and add corrects results to help the neural network to accumulate positive sample cases. We used *FakeNewsNet* dataset to demonstrate the superior accuracy of our method over other state-of-the-art supervised learning methods models. When we used 80% of labeled data for training and 20% of unlabeled data for testing, the self-learning semi-supervised deep learning model based on L achieved a precision of 0.90, a recall score of 0.86, and a F1-score of 0.88, respectively, demonstrating the best performance, about 30% better than those of the machine learning methods. When we used 50% of labeled data for training and 50% of unlabeled data for testing, the precision of our method was 0.88, about 30% higher than that of machine learning methods and 10% higher than that of deep learning methods. When we used 20% of labeled data for training and 80% of unlabeled data for testing, the precision of our method was 0.88, about 40% higher than that of machine learning methods and 15% higher than that of deep learning methods. Which proved that our method performed better and more consistently in the case of incomplete annotated training datasets or relatively small datasets.

In the future work, as the self-learning semi-supervised deep learning network proposed in our paper can automatically return and add correct results with a small amount of labeled data to accumulate positive sample cases, we will collect and establish fake news datasets on social media related to COVID-19, and use semi-supervised learning methods to detect fake news about COVID-19. We will also try to apply self-learning semi-supervised deep learning networks to the detection of multi-source and multi-class fake news.

Funding This research was funded by National Natural Science Foundation of China (61772375, 61936013, 71921002), The National Social Science Fund of China (18ZDA325), National Key R&D Program of China (2019YFC0120003), Natural Science Foundation of Hubei Province of China (2019CFA025); and Independent Research Project of School of Information Management Wuhan University (413100032).

The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Abbreviation *COVID-19,* The New Coronavirus Disease 2019; *SVM,* Support Vector Machines; *LR,* Logistic Regression; *NB,* Naive Bayes; *CNN,* Convolutional Neural Network; *(BERT) model,* The Bidirectional Encoder Representations from Transformers model; *TI-CNN,* Text and Image information based Convolutinal Neural Network; *RNN,* Recurrent Neural Networks; *LSTM,* Long Short-term Memory; *BI-LSTM,* Bidirectional Long Short-term Memory; *TP,* True Positive; *FP,* False Positive; *FN,* False Negative

References

1. Ahmed H, Traore I, Saad S (2017) Detection of online fake news using N-gram analysis and machine learning techniques. Lect Notes Comput Sci 10618:127–138
2. Boididou C, Middleton SE, Jin Z, Papadopoulos S, Dang-Nguyen DT, Boato G, Kompatsiaris Y (2018) Verifying information with multimedia content on twitter. Multimed Tools Appl 77:15545–15571
3. Granik M, Mesyura V (2017) Fake news detection using naive Bayes classifier[C]. IEEE First Ukraine Conference on Electrical & Computer Engineering
4. Huang YF, Chen PH (2020) Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms[J]. Expert Systems with Applications 159:113584
5. Julio C, Reis S, Correia A, Murai F (2019) Explainable Machine Learning for Fake News Detection. Proceedings of the 10th ACM Conference on Web Science, 2019, pp 17–26

6. Jwa H, Oh D, Park K, Kang J, Lim H (2019) Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT) [J]. Appl Sci 9(19):4062
7. Mitra T (2015) CREDBANK: a large-scale social media corpus with associated credibility annotations[J]. In ICWSM'15
8. Okoro EM, Abara BA, Umagba AO, Ajonye AA, Isa ZS (2018) A hybrid approach to fake news detection on social media. Niger J Technol 37(2):454–462
9. Papanastasiou Y (2017) Fake news propagation and detection: a sequential model[J]. SSRN Electron J
10. Rashed KAN, Renzel D, Klamma R, Jarke M (2014) Community and trust-aware fake media detection. Multimed Tools Appl 70:1069–1098
11. Ruchansky N, Seo S, Liu Y (2017) CSI: a hybrid deep model for fake news detection[J]. The 2017 ACM Conference
12. Santia GC, Williams JR (2018) Buzzface: A news veracity dataset with facebook user commentary and egos. In ICWSM'18
13. Shu K, Sliva A, Wang S et al (2017) Fake news detection on social media: a data mining perspective[J]. ACM SIGKDD Explorations Newsletter 19(1)
14. Shu K, Wang S, Liu H (2017) Exploiting tri-relationship for fake news detection. arXiv preprint arXiv: 1712.07709.
15. Shu K, Mahudeswaran D, Wang S et al (2018) FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. arXiv preprint arXiv: 1809.01286
16. Singh L, Bansal S, Bode L et al (2020) A first look at COVID-19 information and misinformation sharing on Twitter[J]. arXiv preprint arXiv:2003.13907
17. Tacchini E, Ballarin G, Della Vedova ML et al (2017) Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506
18. Wang W Y (2017) "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection[J]. arXiv preprint arXiv:1705.00648
19. Wang Y, Ma F, Jin Z et al (2018) EANN: Event adversarial neural networks for multi-modal fake news detection. In KDD'18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 19-23, 2018, London, UK. https://doi.org/10.1145/3219819.3219903
20. Yang Y, Zheng L, Zhang J et al (2018) TI-CNN: Convolutional Neural Networks for Fake News Detection[J]. arXiv preprint arXiv: 1806.00749
21. Zhang J, Dong B, Yu PS (2018) Fake news detection with deep diffusive network model[J]. arXiv preprint arXiv:1805.08751