Information Processing and Management 58 (2021) 102418

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Convolutional neural network with margin loss for fake news detection

Mohammad Hadi Goldani, Reza Safabakhsh, Saeedeh Momtazi *

Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO

Keywords:
Fake news detection
Convolutional neural networks
Non-static word embedding
Margin loss function

ABSTRACT

The advent of online news platforms such as social media, news blogs, and online newspapers in recent years and their facilitated features such as swift information flow, easy access, and low costs encourage people to seek and raise their information by consuming their provided news. Furthermore, these platforms increase the opportunities for deceiver parties to influence public opinion and awareness by producing fake news, i.e., the news which consists of false and deceptive information and is published for achieving specific political and economic gains. Since the discerning of fake news through their contents by individuals is very difficult, the existence of an automatic fake news detection approach for preventing the spread of such false information is mandatory. In this paper, Convolutional Neural Networks (CNN) with margin loss and different embedding models proposed for detecting fake news. We compare static word embeddings with the non-static embeddings that provide the possibility of incrementally up-training and updating word embedding in the training phase. Our proposed architectures are evaluated on two recent well-known datasets in the field, namely ISOT and LIAR. Our results on the best architecture show encouraging performance, outperforming the state-of-the-art methods by 7.9% on ISOT and 2.1% on the test set of the LIAR dataset.

1. Introduction

Nowadays, because of the widespread usage of social media, we are witnessing enormous consequences for society, business, and culture that have this potential to be negative and positive. As a positive effect, it can help to control the crisis faster (Bondielli & Marcelloni, 2019). On the other hand, as a negative effect, people manipulate real information due to political, economic, or social motivations. The spread of this misleading information can be harmful.

News, as a type of information, has more potential to use for misleading information. For example, through tweets, people share articles, photos, and videos such that almost 85% of the topics discussed on Twitter relate to the news (Smart Insights, 2019). Moreover, in social media, the freedom of a user to post anything results in the spread of false information. When this false information is presented as news statements, it is called fake news, which can be a type of propaganda or yellow journalism that consists of misinformation.

People usually spread news shared by their friends more quickly without any validation/verification. A fake news detection system aims to help users detect and filter out potentially deceptive news. The prediction of intentionally misleading news is based on the analysis of truthful and fraudulent of previously reviewed news.

* Corresponding author.
E-mail addresses: goldani@aut.ac.ir (M.H. Goldani), safa@aut.ac.ir (R. Safabakhsh), momtazi@aut.ac.ir (S. Momtazi).

https://doi.org/10.1016/j.ipm.2020.102418
Received 31 January 2020; Received in revised form 2 October 2020; Accepted 19 October 2020
Available online 29 October 2020

Fake news has traditionally been spread through print and broadcast mediums, but with the rise of social media, it can now be disseminated virally (Thota, Tilak, Ahluwalia, & Lohia, 2018). The task of fake news detection can be a simple binary classification, or in a multi-label dataset, can be a fine-grained classification (Tin, 2018). After the introduction of public datasets, such as the Kaggle dataset, the LIAR dataset, and the ISOT dataset, researchers tried to increase the performance of their models using these datasets (Meel & Vishwakarma, 2019).

So far, attempts for detecting fake news are limited to some conventional machine learning methods. The performance of neural networks for fake news detection can be improved by using different settings and components. one of the main settings in the training of a model is the loss function that, in recent years, has been used to improve the performance in many classification tasks(Li, Yu, Chang, Ma, & Cao, 2019).

In this paper, we propose a new model based on CNNs for detecting fake news. In our proposed model, the CNN architecture for detecting fake news is enhanced by using margin loss. We also compare different varieties of word embeddings. We show proposed models achieve better results in comparison to the state-of-the-art methods.

The rest of the paper is organized as follows: Section 2 reviews related works about fake news detection. Section 3 presents the model proposed in this paper. The datasets used for fake news detection and evaluation metrics are introduced in Section 4. Section 5 reports the experimental results, comparison with the baseline classification, and discussion. Finally, Section 6 summarizes the paper and concludes the work.

2. Related work

Social media have become the main source of information. Distinguishing rumors from the truth in this huge volume of information is very crucial and difficult. Misinformation identification has been studied widely in recent years, resulting in the introduction of different tasks in this field. In addition to the fake new detection task, which is the focus of this paper, the task of detecting unverified information has been studied as rumor veracity identification and stance detection. According to the nature of these tasks, different datasets have been presented to simulate unverified information in social media. Considering the difference between the tasks, the datasets were provided in different settings and frameworks which have been used for evaluating different machine learning and deep learning methods proposed by researchers.

For rumor veracity identification, Zhang, Lipani, Liang, and Yilmaz (2019) proposed a Bayesian deep learning model to address the problem of poor performance in representing the uncertainty of the prediction. Their model at first encodes a claim to be verified and generates a prior belief distribution; then, in order to summarize the content, the model encodes all replies to the claim in their temporal order through an LSTM. This summary is then used for generating the posterior belief by updating the prior belief. For training the model, they develop a stochastic gradient variational Bayesian algorithm to approximate the analytically intractable posterior distribution. They showed that their model outperforms the state-of-the-art methods.

To distinguish rumor and non-rumor tweets with both linguistic and user-based features, Singh, Kumar, Rana, and Dwivedi (2020) proposed an attention-based Long-Short Term Memory (LSTM) network which that uses tweet text. They showed the superiority of their model compared to the state-of-the-art methods that used conventional machine and deep learning models for rumor detection.

Alkhodair, Ding, Fung, and Liu (2020) investigated the problem of detecting breaking news rumors, rather than long-lasting rumors, that spread in social media. They propose a model to mitigate the topic shift issues. The model automatically identifies rumors, learns word embeddings, and trains a Recurrent Neural Network (RNN) with two different objectives. They used a real-life rumor dataset. Their experiment simulated a cross-topic emerging rumor detection scenario. They showed the superiority of their model compared to the state-of-the-art methods for rumor detection.

For stance detection, which is another main task in the field, Sobhani, Inkpen, and Zhu (2019) proposed a deep learning model for multi-target stance detection. They claimed that the previous works often treated each target independently and ignored the potential dependency among targets. They showed that an attention-based encoder-decoder framework is very effective for solving this problem and improving the performance of methods that jointly learn dependent subjectivity through cascading classification.

Zhang, Liang, Lipani, Ren, and Yilmaz (2019) proposed a model to overcome one of the major problems facing the current machine learning models used for stance detection. This problem is a severe class imbalance among the four classes, agree, disagree, discuss, and unrelated, of a stance detection task. In this model, they propose a hierarchical representation of these classes, which combines four classes under a new related class. In addition, they used a two-layer neural network that learns from this new representation and controls the error propagation between the two layers by a maximum mean discrepancy regularizer. They showed the superiority of their model in terms of accuracy compared to the state-of-the-art methods for stance detection. Also, respect to controversial topics for detecting the stance of prolific Twitter users, Darwish, Stefanov, Aupetit, and Nakov (2020) proposed a new method. In this method, dimensionality reduction is used to project users into a low-dimensional space, and clustering is used to find core users who are representative of the different stances. They claimed that their framework has three major advantages compared to the past methods, which are based on semi-supervised or supervised classification: (1)the method creates clusters without requiring any prior labeling of the users; (2)to conduct the actual labeling or to specify the relevant stances (labels), they need neither the domain- nor topic-level knowledge; (3)Third, in the face of data skewness, their framework is robust.

As mentioned, our focus is on fake news detection, which is mainly based on supervised methods such that a machine/deep learning model is built on available fake news labeled data are containing both fake and real news. The model is then used to make the decision on new incoming news to find out if it is fake or not. Most of the available works that investigated the task of fake news detection are done in recent years, which will be described in this section. Moreover, in line with these efforts, some binary or multi-class datasets are introduced as resources of the field. Ahmed, Traore, and Saad (2017) introduced a new binary class dataset, called ISOT, that was

collected from the real-world sources for fake news detection. In addition, two new datasets with seven different domains were proposed by Pérez-Rosas, Kleinberg, Lefevre, and Mihalcea (2018). Their datasets include actual news excerpts and do not contain any short statements of fake news information. Furthermore, Wang (2017) introduced a new multi-class dataset, called LIAR, including 12, 836 labeled short statements. This dataset was collected from more natural contexts such as tweets, Facebook posts, political debates, etc.

2.1. Fake news detection with conventional machine learning models

Many researches based on machine learning presented in the task of fake news detection. An overview of deception assessment approaches are proposed by Conroy, Rubin, and Chen (2015) that consist of the final goals of these approaches and the major classes. In this work, two approaches were used for investigating the problem: (1) linguistic methods that extract language patterns and analyzed news content; (2)network-based methods. In this section, for decision making about new incoming news, network queries and massage metadata were deployed.

As another work, based on the users who interact with the news, another approach was proposed that can predict if the news is fake or not. According to this approach, two classification methods are used: (1) logistic regression model that incorporates the user interaction as the features, and (2) the novel adaptation of the boolean label crowdsourcing techniques. Based on the reported results, both methods achieved high accuracy. Also, the authors proved that the information of users who interact with the news is one of the important features that can be considered for fake news detection(Tacchini, Ballarin, Vedova, Moret, & de Alfaro, 2017). An automated fake news detector, namely CSI, proposed by Ruchansky, Seo, and Liu (2017) that includes three modules: capture, score, and integrate. For prediction, this detector uses text, response, and source of the news. This model extracts the temporal representations of news articles. The model then represents and scores the behavior of the users. The outputs of these two modules have been used for the classification of both users and articles. They showed that their model improves the accuracy of fake news detection.

After proposing the ISOT dataset, The authors evaluated six machine learning techniques and n-gram models on the ISOT dataset. They showed that TF-IDF as the feature extractor and linear support vector machine as the classifier achieves the best performance (Ahmed et al., 2017). After proposing two new datasets by Pérez-Rosas et al. (2018), they also showed by using linguistic features such as lexical, syntactic, and semantic level features and a linear support vector machine classifier, fake and real news can be detected. They also showed that the performance of their system is comparable to that of humans in this area. Another algorithm, called Detective, was introduced by Tschiatschek, Singla, Rodriguez, Merchant, and Krause (2018). This algorithm is a tractable Bayesian algorithm that tries to provide a balance between selecting news that directly maximizes the objective value and selecting that aids toward learning the user's flagging accuracy. The primary goal of their research was minimizing and reducing the number of users who have seen the fake news before it becomes blocked. They showed that their proposed model is competitive with the fictitious algorithm OPT which knows the true users' parameters. They also show that the Detective algorithm is robust in applying flags even in a setting where the majority of users are adversarial.

2.2. Fake news detection with deep learning models

Many approaches based on neural networks and deep learning models used for detecting fake news articles. Some of the models used CNN; some used RNN, and others used hybrid models or more recent neural networks like Capsule neural networks.Wang (2017) proposed the LIAR dataset. Then he presented a model that used statements and meta-data together as input and a CNN for feature extraction from statements and a Bi-directional Long Short Term Memory (BiLSTM) network for feature extraction from the meta-data. They showed that their model achieved significant improvements in terms of accuracy. A model that incorporates speaker profiles as features proposed by Long, Lu, Xiang, Li, and Huang (2017) on the LIAR dataset, that contains party affiliation, speaker location, title, and credit history, into an attention-based LSTM model. They used two different ways for contributing speaker profiles to the model; (1)considering them in the attention model; (2)including them as additional input data. They showed this model improves the performance of the classifier on the LIAR dataset. The event adversarial neural network was proposed by Wang et al. (2018). This model contains three main components: (1) the multi-modal feature extractor that uses CNN as its core module, (2) the fake news detector that is a fully connected layer with softmax activation which is deployed to predict if a news is fake or not, (3) the event discriminator that two fully connected layers are used and aim at classifying the news into one of K events based on the first component representations.

Also, a model based on Capsule Neural Network (CapsNet) proposed for fake news detection by Goldani, Momtazi, and Safabakhsh (2020). They applied different levels of n-grams and different embedding models for news items with various lengths. For long news statements, four different filter sizes, namely 2,3,4, and 5, were used as n-gram convolutional layers with non-static embedding. For short news statements, only two filters with 3 and 5 were used as kernel size with static embedding. They showed the superiority of their model compared to the state-of-the-art methods in terms of accuracy.

3. CNN with margin loss for fake news detection

In recent years, CNNs used in many computer vision tasks and improved the state-of-the-art performance of many visual classification tasks, such as image classification (Gong, Zhong, Yu, Hu, & Li, 2019), face verification (Amato et al., 2019) and object recognition (Gandarias, Garcia-Cerezo, & Gomez-de Gabriel, 2019). Moreover, it is used in many natural language processing tasks, including text classification (Wang, Huang, & Deng, 2018; Yao, Mao, & Luo, 2019), aspect extraction (Xu, Liu, Shu, & Philip, 2018),

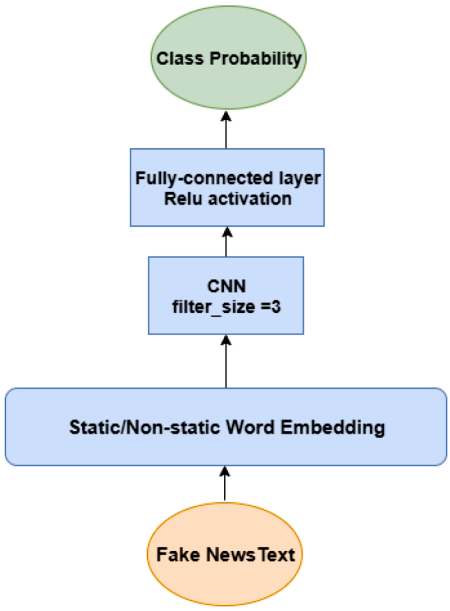


Fig. 1. Proposed model for fake news detection.

and sentiment classification (Huang & Carley, 2018). The architecture of CNNs with convolution and pooling layers can carefully extract features from local to global renders that is a strong representation ability for CNNs.

One of the most commonly used loss functions that are used with softmax in CNNs is cross-entropy. The advantages of CNN with cross-entropy are its simplicity and excellent performance, but as a negative point, this network does not explicitly encourage discriminative learning of features (Liu, Wen, Yu, & Yang, 2016). To extract stronger features for the learning process, CNNs should be reinforced with more discriminative information. For this goal, the intra-cluster similarity and inter-cluster dissimilarity of the learned features should be maximized. As a solution, other loss functions such as contrastive loss (Hadsell, Chopra, & LeCun, 2006) and triplet loss (Schroff, Kalenichenko, & Philbin, 2015) were proposed. These loss functions enforce extra intra-cluster compactness and inter-cluster separability. In this paper, we use the margin loss for CNNs, which is inspired by the loss function proposed by Sabour, Frosst, and Hinton (2017) in capsule neural networks for image classification.

As mentioned, instead of using the cross-entropy loss, we propose using margin loss within the CNN architecture.

Sabour et al. (2017) applied a fixed margin loss for the classification of digit images. They use margin-loss to classify a multi-label dataset better, and the use of this loss function can avoid the overlapping problem and help the model to alleviate the problem of overfitting. They set the margin empirically. Eq. (1) shows the margin loss where *y* is true-label vector, *predict* is the predicted values, and *lambda* is a parameter that is empirically set.

$$L = \left[y * \max(0, (0.9 - predict))^2\right] + \left[\lambda * (1 - y) * \max(0, (predict - 0.1))^2\right] \quad (1)$$

We use this idea and propose a model based on CNN with the margin loss, as presented in Fig. 1. It consists of three-layers: word embedding layer, CNN layer, and fully connected layer. Assuming that the input is a news text, the goal of this model is to predict the class probability. For binary classification, we have two classes, real and fake, and for a multi-label data set, we have more than two classes.

3.1. Embedding layer

In this layer, we use 'Glove.6B.300d' as a pre-trained word embedding. Moreover, we incorporate different types of word embedding models for fake news detection. These models are a group of highly efficient computational models that are created by training neural networks with two layers trained on a large volume of text. The output vector representations are several hundred dimensions for every word. In these representations, words with similar meanings are mapped close to each other in the vector space (Mikolov, Chen, Corrado, & Dean, 2013).

One of the common methods to improve text processing performance, especially in the absence of a large supervised training set, is using pre-trained vectors for initializing word vectors. Pre-trained embeddings are produced by training on a large volume of text. These distributed vectors are fed into deep neural networks and used for many text classification tasks (Kim, 2014).

By applying different learning settings for vector representation of words via word2vec, Kim (2014) showed their superiority compared to the regular pre-trained embeddings when they are used within a CNN model. These settings are as follow:

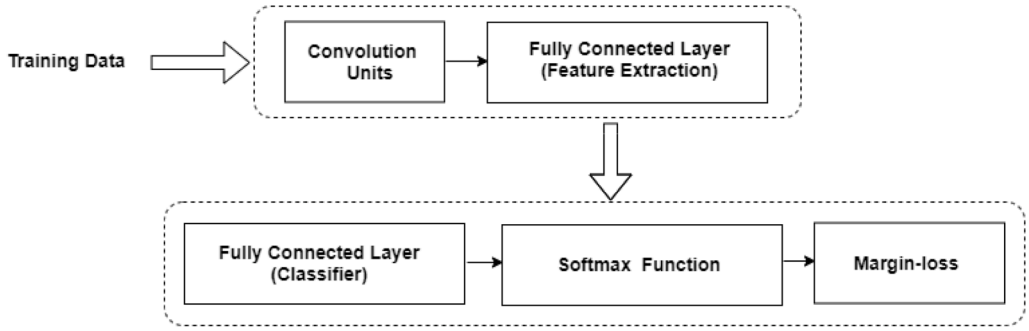


Fig. 2. The architecture of CNN with margin loss.

Table 1
The LIAR dataset statistics provided by Wang (2017).

LIAR Dataset Statistics	
Training set size	10,269
Validation set size	1,284
Testing set size	1,283
Avg. statement length (tokens)	17.9
Top-3 Speaker Affiliations	
Democrats	4,150
Republicans	5,687
None (e.g., FB posts)	2,185

- **Static word2vec model:** in this model, as input to the neural network architecture, pre-trained vectors are used. During training, these vectors are kept static, and only the other parameters are learned.
- **Non-static word2vec model:** in this model, at the initialization of learning, pre-trained vectors are used, but these vectors are fine-tuned during the training phase of each task using the training data of the target task.
- **Multichannel word2vec model:** in this model, both of the previous settings are used for every half part of vectors, i.e., during the training, one part is static, and another part is fine-tuned.

3.2. CNN layer

In this layer, a convolutional layer is used, which extracts 3-gram features of a sentence through a convolutional filter with size 3.

In this part, the input matrix consists of the news; each row represents a word of the text in the embedding space. The filter then tries to capture sequences of words. The filter performs convolutions on the sentence matrix and generates feature maps; then, max pooling is performed over each map.

3.3. Fully connected layer

The fully connected layer has been used in two-parts of the architecture. The first part is following the CNN unit. In this part, the fully connected layer acts as a feature extractor. In the second part, another fully connected layer is used for the classification purpose. In this part, the softmax function and margin loss are utilized in our model. Fig. 2 shows the structure of our proposed model.

4. Evaluation

4.1. Dataset

For evaluation of the proposed model, we use two datasets, namely LIAR (Wang, 2017), and ISOT fake news (Ahmed et al., 2017). In this section, these two datasets are introduced.

4.1.1. The LIAR dataset

The LIAR dataset is one of the recent well-known datasets that is provided by Wang (2017). This dataset was collected from `POLITIFACT.COM` API and included 12.8K human-labeled short statements. For validation of this dataset, `POLITIFACT.COM` editor is used. For the degree of truthfulness, six fine-grained labels, namely true, false, barely-true, half-true, mostly-true, and pants-fire, are considered. For each news item, in addition to news statements, metadata as speaker profiles are considered. These metadata include valuable information about the speaker's name, subject, job, state, party, and total credit history counts of the new speaker. The total

Table 2
Type and size of articles per category for ISOT dataset provided by Ahmed et al. (2017).

News Type	Total size	Subject	
		Type	Size
Real-News	21417	World-News	10,145
		Politics-News	11,272
Fake-News	23481	Government-News	1570
		Middle-east	778
		US News	783
		Left-News	4459
		Politics	6841
		News	9050

credit history counts include false counts, barely-true counts, mostly-true counts, half-true counts, and pants-fire counts. Table 1 shows the statistics of the LIAR dataset.

4.1.2. The ISOT fake news dataset

As mentioned in Section 2, ISOT dataset was provided by Ahmed et al. (2017) from real-world sources[1]. This dataset was collected from `Reuters.com` and `Kaggle.com`. In the ISOT dataset, every instance is longer than 200 characters and in addition to news statements several metadata such as article type, article text, article title, and article date are available. The statistics of the ISOT dataset are shown in Table 2.

4.2. Experimental setup

The experiments of this paper were conducted on a PC with Intel Core i7 6700k, 3.40GHz CPU; 16GB RAM; Nvidia GeForce GTX 1080Ti GPU in a Linux workstation. For implementing the proposed model, the Keras library (Chollet et al., 2015) was used, which is a high-level neural network API. In the experiments, we use the word2vec tool to get word embeddings. To initialize the embedding layer for the LIAR and ISOT datasets, we use the 300-dimensional word vectors that are pre-trained in the Glove setting (Pennington, Socher, & Manning, 2014). Additionally, we use non-static and multi-channel embedding mentioned in Section 3. For both datasets we use the Adam optimizer(Kingma & Ba, 2015) and Relu activation function. The hyperparameter settings for the LIAR dataset are 150 for batch size, 0.25 for lambda, and 0.0005 for the learning rate. For the ISOT dataset, we choose 25 for the batch size, 0.7 for lambda, and 0.001 for the learning rate.

4.3. Evaluation metrics

The evaluation metric in our experiments is the classification accuracy. Accuracy is the ratio of correct predictions to the total number of samples and is computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{2}$$

where TP represents the number of True Positive results, FP represents the number of False Positive results, TN represents the number of True Negative results, and FN represents the number of False Negative results.

5. Results

For evaluating the effectiveness of the proposed model on two datasets, a series of experiments were performed. In this section, the experiments are explained, and the results are compared to other baseline methods. Moreover, the results for every dataset are discussed separately.

5.1. Classification for the LIAR dataset

As mentioned in Section 4.1.1, the LIAR dataset is a multi-label dataset with short news statements. In comparison to the binary classification task for fake news detection, classification on this dataset is more challenging. In this dataset, the train set, the validation set, and the test set were separated according to Wang (2017). For the evaluation of the proposed model, we use speaker profiles as metadata. The accuracy of the proposed model on both the validation set and test set is computed. The results of the proposed model are compared with the state-of-the-art baselines including hybrid CNN (Wang, 2017), LSTM with attention (Long et al., 2017), and CapsNet (Goldani et al., 2020). Table 3 shows these results for the test set.

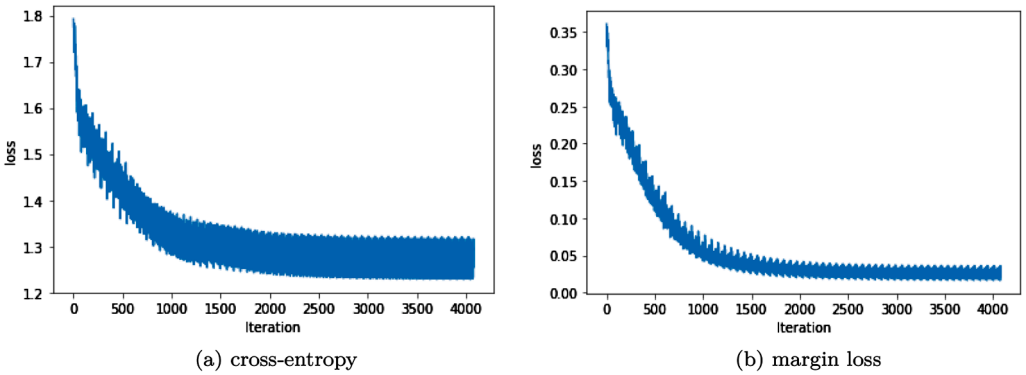
[1] http://www.uvic.ca/engineering/ece/isot/datasets/index.php

Table 3
Comparison of CNN with margin loss(CNN-ML) result with Hybrid CNN(H-CNN), LSTM with attention(A-LSTM), CapsNet, and CNN with cross-entropy(CNN-CE) baseline on test set.

Data	Meta-data	H-CNN	LSTM-A	CapsNet	CNN-CE	CNN-ML
Text stm	Party	24.8	**25.7**	24	23.1	24.2
	State	25.6	**26.8**	24.3	22.6	25
	Job	**25.8**	25.7	25.1	24.5	25.7
	History	24.1	38.5	39.5	37.1	**41.6**

Table 4
The results of proposed model with different word embeddings on LIAR dataset.

Model	Embedding type	Valid	Test
CNN with margin loss	Static	**44.4**	**41.6**
	Non-static	43.8	40.0
	Multi-channel	43.3	40.1



(a) cross-entropy (b) margin loss

Fig. 3. training loss for different loss functions on LIAR dataset.

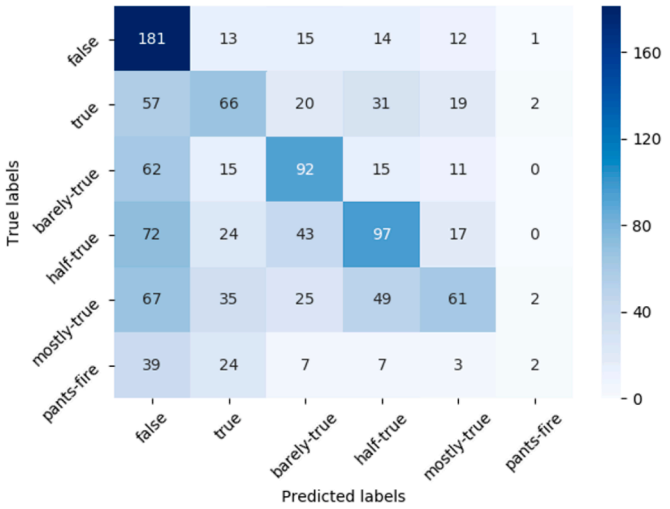


Fig. 4. Confusion matrix of classification using proposed model for LIAR dataset.

The best result of the model for the test set is achieved by using history as metadata. The results show that this model can perform better than CNN with cross-entropy and state-of-the-art baselines including hybrid CNN (Wang, 2017), LSTM with attention (Long et al., 2017), and CapsNet (Goldani et al., 2020) by 2.1% on the test set.

As another experiment, the proposed model is evaluated with different word embeddings as described in Section 3. Table 4 shows

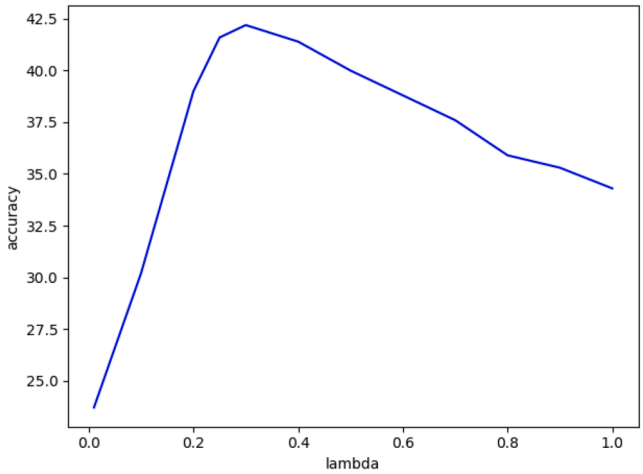


Fig. 5. accuracy on testset of LIAR dataset for different values of lambda.

Table 5
Three samples with wrong prediction on LIAR data set.

Statement	Predicted	True label
Says the unemployment rate for college graduates is 4.4 percent and over 10 percent for noncollege-educated.	half-true	true
The Fed created $1.2 trillion out of nothing, gave it to banks, and some of them foreign banks, so that they could stabilize their operations.	mostly-true	true
Says he won the second debate with Hillary Clinton in a landslide in every poll.	false	pants-fire

Table 6
Comparison of the proposed CNN with margin loss with the results reported by Ahmed et al. (2017).

Model	Meta-data	Test Accuracy (%)
SVM	Article Text	86
LSVM	Article Text	92
KNN	Article Text	83
Decision Tree	Article Text	89
SGD	Article Text	89
Linear Regression	Article Text	89
CNN with CE	Article Text	97
CNN with margin loss	Article Text	**99.1**

the result of applying different word embeddings for the proposed model on the validation set and test set of the LIAR datasets by using history as metadata. The best result is achieved by applying static word embedding. These results show that fine-tuning the embeddings using the training data cannot improve the performance of the model. One potential reason is that the size of the dataset is not too large (12.8K) to be able to enhance the quality of embeddings.

5.1.1. Discussion

Fig. 3 shows the training loss for CNN with cross-entropy and margin loss. In comparison to the margin loss, cross-entropy leads to significantly higher loss values and higher oscillations as network training approach convergence.

Fig. 4 shows the confusion matrix of the best performance of the proposed model for the test set. The model classifies false, barely-true, and half-true news with more accuracy. Nevertheless, it is difficult to distinguish between true and half-true and also between barely-true and false. Classifying pants-fire is more challenging, and a large number of texts from pants-fire are predicted as false.

As mentioned in Section 3, the parameter of lambda in margin loss is empirically set. As another experiment, we test different values of lambda between 0 and 1 on the test set of the LIAR dataset. Fig. 5 shows accuracy on the test set of the LIAR dataset for different values of lambda when history is used for metadata. For this dataset, 0.25 set for lambda.

Table 5 shows the statements of three wrongly predicted samples for detecting fake news on the LIAR dataset. For analyzing the results, we investigate the effect of words in training statements that are tagged incorrectly.

By analyzing the statements, the main words that cause the wrong prediction are negative words, in the first statement, *un* and *non*, for the second statement, *nothing* and the last one *every* has this potential to considered false.

Table 7
Comparison of CNN with margin loss function for different word embedding types for ISOT dataset.

Model	Embedding type	Results
CNN with margin loss	Static	99.1
	Non-static	**99.9**
	Multi-channel	99.8

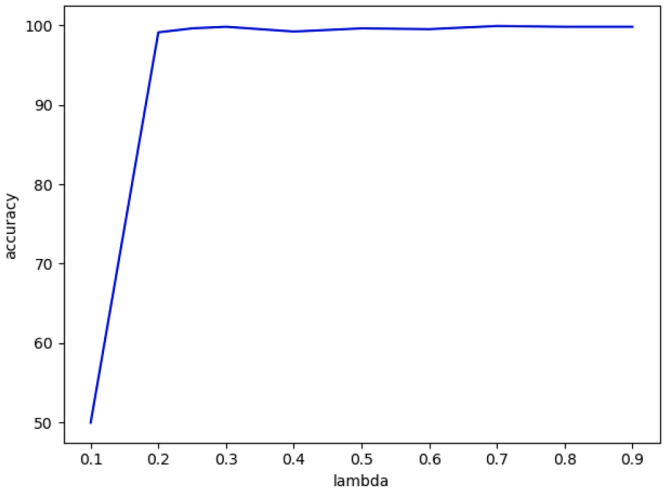


Fig. 6. accuracy on ISOT dataset for different values of lambda.

Table 8
Two samples with wrong prediction.

Statement title	Predicted	True label
Factbox: What's in tax bill from Trump, House Republicans?	fake	real
Trump Just Got His P*ssy Handed To Him By New Zealand's Female Prime Minister	real	fake

5.2. Classification for ISOT dataset

As mentioned in Section 4, Ahmed et al. (2017) presented the ISOT dataset. For evaluating the proposed model on the ISOT dataset according to the baseline paper, 1000 articles for every set of real and fake articles, a total of 2000 articles, are considered as a test set, and the proposed model is trained with the rest of the data.

Different machine learning methods are evaluated for fake news detection on the ISOT dataset by Ahmed et al. (2017). These approaches include the Support Vector Machine (SVM), the Linear Support Vector Machine (LSVM), the K-Nearest Neighbor (KNN), the Decision Tree (DT), the Stochastic Gradient Descent (SGD), and the Logistic regression (LR). Table 6 shows the performance of the proposed CNN with margin loss for fake news detection in comparison to other methods. The accuracy of our model is 7.1% higher than the best result achieved by LSVM.

In the next step, the proposed model is evaluated with different word embeddings. The results of applying different word embeddings are shown in Table 7. In contrast to the results on LIAR, the best result on ISOT is achieved by applying non-static embedding. This shows that using large training data (45k), we can further improve the performance when fine-tuning the embeddings for the target task.

As another experiment, different values of lambda between 0 and 1 on the test set of the ISOT dataset are tested. Fig. 6 shows accuracy on the test set of ISOT dataset for different values of lambda. For this dataset, 0.7 is set for lambda.

5.2.1. Discussion

As mentioned in Section 5.2, the accuracy of predicting true labels by the proposed model is very high, and there is a very small number of wrong predictions. Table 8 shows the title of two wrongly predicted samples for detecting fake news on the ISOT dataset. For analyzing the results, we investigate the effect of sample words frequency in training statements that are tagged as real and fake separately.

For this work, all of the words and their frequencies from the two wrong samples and both real and fake labels of the training data are extracted. Table 9 shows the statistics of this data. Then stop words for every sample are omitted. After that, four more frequent

Table 9
The number of word tokens and word types of training data and samples.

Data	Word tokens	Word types
Training data With real label	8,264,220	76,213
Training data With fake label	10,115,367	92,613
Sample 1 (Predicted fake but is real)	458	199
Sample 2 (Predicted real but is fake)	202	134

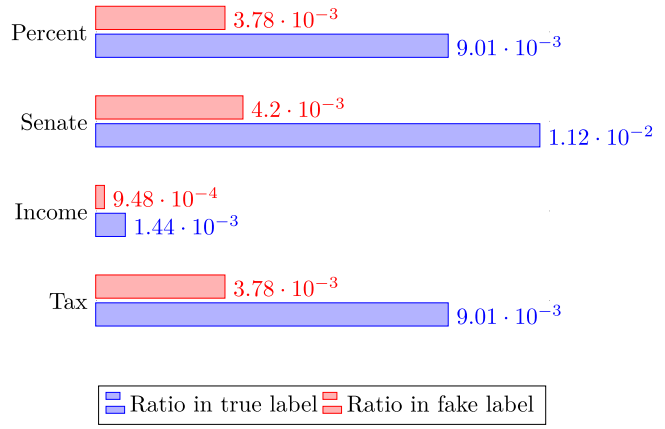


Fig. 7. Normalized frequency for words in sample 1 and training data with fake and real label.

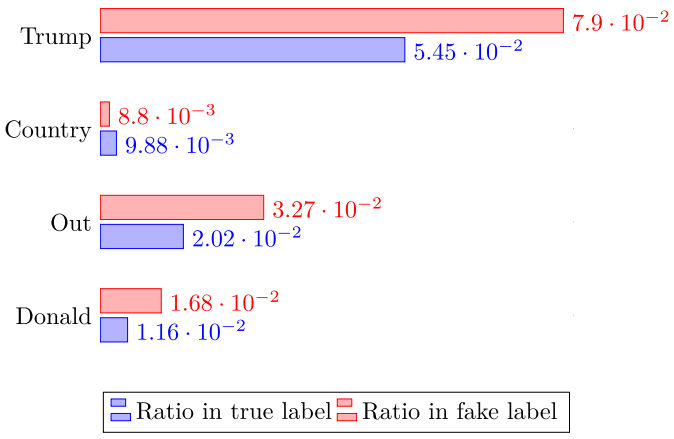


Fig. 8. Normalized frequency for words in sample 2 and training data with fake and real label.

words for every sample are chosen. In this part, the frequencies of these words are normalized. For ease of comparison, the normalized frequencies of words in real and fake labels of training data and samples are multiplied by 10. This information for every sample is demonstrated in Figs. 7 and 8.

The text of Sample 1 is predicted as fake news, but it is real. Fig. 8 shows four frequent words of Sample 1, namely "*Percent*", "*Senate*", "*Income*", and "*Tax*". The frequency of these words in training data with real labels is more than the frequency of these words in training data with fake labels.

The label of Sample 2 is predicted as real, but it is fake. In Fig. 8, the four most frequent words of Sample 2, namely "*Trump*", "*country*", "*out*", and "*Donald*", are considered. Obviously, the frequency of these words in training data with fake labels is more. This fact shows the strong effect of the frequency of the sample words on the prediction of the labels.

6. Conclusion

In this paper, we propose CNN with margin loss for fake news detection. Following the success history of margin loss in image classification, to the best knowledge of the authors, this is the first time that this idea is used for a text classification task. Moreover, we incorporate different embedding models and show that applying the non-static embedding model, which incrementally up-trains and updates the word embeddings in the training phase, can improve the performance of the proposed model when the size of training data is large enough. For the evaluation of the proposed model, two well-known datasets, namely ISOT and LIAR, are used. We use static word embedding for the LIAR dataset and non-static word embedding for the ISOT. The experimental results on these two datasets showed improvement in terms of accuracy by 7.9% on the ISOT dataset and 2.1% on the test set of the LIAR dataset.

CRediT authorship contribution statement

Mohammad Hadi Goldani: Conceptualization, Formal analysis, Writing - original draft. **Reza Safabakhsh:** Conceptualization, Supervision, Writing - review & editing. **Saeedeh Momtazi:** Conceptualization, Supervision, Writing - review & editing.

References

Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using N-gram analysis and machine learning techniques. *International conference on intelligent, secure, and dependable systems in distributed and cloud environments* (pp. 127–138). Springer.

Alkhodair, S. A., Ding, S. H., Fung, B. C., & Liu, J. (2020). Detecting breaking news rumors of emerging topics in social media. *Information Processing & Management, 57* (2), 102018.

Amato, G., Falchi, F., Gennaro, C., Massoli, F. V., Passalis, N., Tefas, A., … Vairo, C. (2019). Face verification and recognition for digital forensics and information security. *2019 7th international symposium on digital forensics and security (ISDFS)* (pp. 1–6). IEEE.

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences, 497*, 38–55.

Chollet, F. et al. (2015). Keras. https://github.com/fchollet/keras.

Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news*, 52. Proceedings of the association for information science and technology* (pp. 1–4).

Darwish, K., Stefanov, P., Aupetit, M., & Nakov, P. (2020). Unsupervised user stance detection on twitter*, 14. Proceedings of the international AAAI conference on web and social media* (pp. 141–152).

Gandarias, J. M., Garcia-Cerezo, A. J., & Gomez-de Gabriel, J. M. (2019). CNN-based methods for object recognition with high-resolution tactile sensors. *IEEE Sensors Journal, 19*(16), 6872–6882.

Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2020). Detecting fake news with capsule neural networks. arXiv preprint arXiv:2002.01030.

Gong, Z., Zhong, P., Yu, Y., Hu, W., & Li, S. (2019). A CNN with multiscale convolution and diversified metric for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 57*(6), 3599–3618.

Hadsell, R., Chopra, S., & LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping*, 2. 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)* (pp. 1735–1742). IEEE.

Huang, B., & Carley, K. M. (2018). Parameterized convolutional neural networks for aspect level sentiment classification. *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 1091–1096).

Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1746–1751).

Kingma, D. P., & Ba, J. (2015). Adam: a method for stochastic optimization. *The international conference on learning representations*.

Li, X., Yu, L., Chang, D., Ma, Z., & Cao, J. (2019). Dual cross-entropy loss for small-sample fine-grained vehicle classification. *IEEE Transactions on Vehicular Technology, 68*(5), 4204–4212.

Liu, W., Wen, Y., Yu, Z., & Yang, M. (2016). Large-margin softmax loss for convolutional neural networks*., 2. ICML* (p. 7).

Long, Y., Lu, Q., Xiang, R., Li, M., & Huang, C.-R. (2017). Fake news detection through multi-perspective speaker profiles. *Proceedings of the eighth international joint conference on natural language processing (volume 2: Short papers)* (pp. 252–256).

Meel, P., & Vishwakarma, D. K. (2019). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, 112986.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). Automatic detection of fake news. *Proceedings of the international conference on computational linguistics,* (pp. 3391–3401).

Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. *Proceedings of the 2017 ACM on conference on information and knowledge management, ACM* (pp. 797–806).

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Proceedings of the international conference on neural information processing systems (NIPS)* (pp. 3859–3869).

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: a unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815–823).

Singh, J. P., Kumar, A., Rana, N. P., & Dwivedi, Y. K. (2020). Attention-based LSTM network for rumor veracity estimation of tweets. *Information Systems Frontiers*, 1–16.

Smart Insights (2019). Global social media research. https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/, Accessed: 2019-12-20.

Sobhani, P., Inkpen, D., & Zhu, X. (2019). Exploring deep neural networks for multi-target stance detection. *Computational Intelligence, 35*(1), 82–97.

Tacchini, E., Ballarin, G., Vedova, M. L. D., Moret, S., & de Alfaro, L. (2017). Some like it hoax: automated fake news detection in social networks. arXiv preprint arXiv:1704.07506.

Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). Fake news detection: A deep learning approach. *SMU Data Science Review, 1*(3), 10.

Tin, P. T. (2018). A study on deep learning for fake news detection,.

Tschiatschek, S., Singla, A., Rodriguez, M. G., Merchant, A., & Krause, A. (2018). Fake news detection in social networks via crowd signals. *Companion proceedings of the web conference* (pp. 517–524).

Wang, S., Huang, M., & Deng, Z. (2018). Densely connected CNN with multi-scale feature attention for text classification.. *IJCAI* (pp. 4468–4474).

Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection. *Proceedings of the 55th annual meeting of the association for computational linguistics* (p. 422??426).

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., … Gao, J. (2018). EANN: Event adversarial neural networks for multi-modal fake news detection. *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, ACM* (pp. 849–857).

Xu, H., Liu, B., Shu, L., & Philip, S. Y. (2018). Double embeddings and CNN-based sequence labeling for aspect extraction. *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 592–598).

Yao, L., Mao, C., & Luo, Y. (2019). Graph convolutional networks for text classification, *33*. *Proceedings of the AAAI conference on artificial intelligence* (pp. 7370–7377).

Zhang, Q., Liang, S., Lipani, A., Ren, Z., & Yilmaz, E. (2019). From stances' imbalance to their hierarchical representation and detection. *The world wide web conference* (pp. 2323–2332).

Zhang, Q., Lipani, A., Liang, S., & Yilmaz, E. (2019). Reply-aided detection of misinformation via bayesian deep learning. *The world wide web conference* (pp. 2333–2343).