

Graph-based Fake News Detection using a Summarization Technique

Gihwan Kim and Youngjoong Ko

Department of Computer Science and Engineering

Sungkyunkwan University

Suwon, Gyeonggi-do, Korea

kimgihwan3364@gmail.com, yjko@skku.edu

Abstract

Nowadays, fake news is spreading in various ways, and this fake information is causing a lot of social damages. Thus the need to detect fake information is increasing to prevent the damages caused by fake news. In this paper, we propose a novel graph-based fake news detection method using a summarization technique that uses only the document internal information. Our proposed method represents the relationship between all sentences using a graph and the reflection rate of contextual information among sentences is computed by using an attention mechanism. In addition, we improve the performance of fake news detection by utilizing summary information as an important subject of the document. The experimental results demonstrate that our method achieves high accuracy, 91.04%, that is 8.85%p better than the previous method.

1 Introduction

Recently, people are easily exposed to large amount of information in various ways with the development of information propagation methods. But some of the large amount of information contains fake information generated for malicious purposes. This fake information is confusing people and causing a lot of social and economic damages. Therefore, the need to detect fake information is increasing to prevent the damages caused by fake news, and it is being researched industrially and academically (Yang et al., 2012; Castillo et al., 2011; Yan et al., 2015).

Prior fake news detection studies have methods of detecting fake news using external information as well as internal information. The method of using internal information detects fake news by analyzing linguistic features in news, such as news content (Levi et al., 2019), writing styles and con-

sistency (Potthast et al., 2018), and relational structure between sentences (Karimi and Tang, 2019) within the news. On the other hand, that of using external information does by analyzing metadata such as aspects of news spreading (Monti et al., 2019) and user profiles of the people spreading the news (Lu and Li, 2020). However, a collection task for external information requires a lot of time and cost, and it is very difficult to identify external information in all documents. In addition, it is more basic and important to understand a document with internal information, and the construction of structural relationships among sentences within a document is one of effective methods for detecting fake news. Therefore, we propose an effective fake news detection method by structuring a context graph for representing the relationships among sentences based on summarization information.

Since all sentences in a document are strongly related to each other, the contextual information with other sentences should be reflected to generate sentence embeddings. In addition, the attention mechanism is exploited for constructing a context graph so that the different relation strength between sentences influences the contextual information of each sentence in the graph. In the context graph, nodes consist of initial sentences embeddings for all the sentences in a document and the weight of an edge is estimated by an attention score between two ended nodes of the edge. We assume that all the nodes are connected because all the sentence in a document is strong related each other. Then a contextualized sentence embedding on each node is generated by reflecting the neighbor's contextual information. The contextualized sentence embedding of a node is computed by the sum of the products of the attention score between the node and its neighbor node and the initial sentence embedding of the neighbor node.

Because the subject of a document is very im-

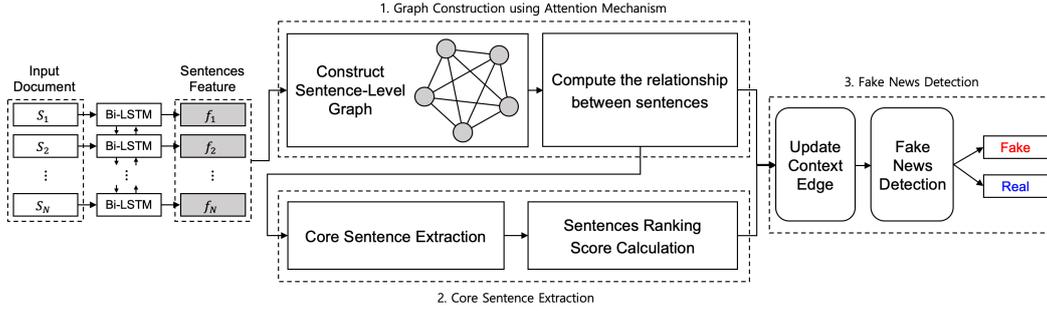


Figure 1: The architecture of our proposed model.

portant information in identifying the content of the document and fake information is commonly related to the subject, we apply a summarization technique (Jeong et al., 2016) that can effectively capture the subject information of the document to fake news detection. By using a summarization technique, sentences containing a lot of subject information are highly ranked and the ranked scores influence the attention scores for construction of the context graph.

For performance comparison with the proposed method, we implemented a baseline model that detects fake news with the sum of all sentence embeddings in a document. Our proposed model shows 91.04% accuracy, that is 11.19%p better performance than baseline model. It also shows 8.85%p better performance compared to other model (Karimi and Tang, 2019) that uses the same dataset and a dependency tree structure among sentences in a document.

2 Notations

We have a corpus D of fake and real news documents. Let a document $d \in D$ contain N sentences s_1, s_2, \dots, s_N and each sentence $s_i \in d$ include words $word_1, word_2, \dots, word_l$ where l denotes the number of words in sentence s_i . We apply the Bi-LSTM network to all sentences of a document and obtain the initial sentence embedding representations $\mathcal{H} = h_1, h_2, \dots, h_N \in \mathbb{R}^{N \times dim}$.

3 Proposed Method

We propose a novel fake news detection method based on the context graph with a summarization technique (see Figure 1). Our method consists of three components. The first one is *graph construction using attention mechanism* for representing the relationship between all sentences using a graph. The second one is *core sentence extraction* for ranking sentences with subject information using a sum-

marization technique. The third one is *fake news detection* for discriminating fake news from a lot of documents.

3.1 Graph Construction using Attention Mechanism

To model the relationship between sentences and their contextual information, we construct a graph $\mathcal{G} = (F, E)$. The graph \mathcal{G} is composed of node (*i.e.*, $F = f_1, f_2, \dots, f_N$) and the edge (*i.e.*, $E = e_{1,2}, e_{ij}, \dots, e_{N-1,N}$). Each node is represented by sentence embedding H and the edge between the i -th node and the j -th node is represented by $e_{i,j}$ and its weight $w_{i,j}$ means the relation strength of the i -th sentence for the j -th sentence. Since all sentences in a document are strongly related to each other, we consider that \mathcal{G} is a fully connected graph. To reflect the different relation strength between sentences, each edge $e_{i,j}$ is associated with a weight $w_{i,j}$, and the weight is derived by the attention mechanism between sentence embeddings h_i and h_j of two nodes f_i and f_j as follows (Eq. 1-3):

$$x_i = ReLU(Wf_i + bias) \quad (1)$$

$$x_j = ReLU(Wf_j + bias) \quad (2)$$

$$w_{i,j} = x_i U x_j \quad (3)$$

where $W \in \mathbb{R}^{m \times dim}$ and $x \in \mathbb{R}^m$. The dimension size of sentence embedding can be reduced to avoid overfitting without weakening the LSTM's capacity (Dozat and Manning, 2018) by Equations (1) and (2). U is a weight matrix to compute the relation strength between x_i and x_j . $w_{i,j}$ represent an edge weight between pair of nodes f_i and f_j in the graph \mathcal{G} .

3.2 Core Sentence Extraction

Based on the constructed graph, the subject information of a sentence is estimated by summing attention scores (edge scores) between the node of

the sentence and its adjacent nodes, and a sentence with the most subject information is extracted as a core sentence. $Core_{sent}$ represents an extracted sentence.

$$Core_{sent} = \operatorname{argmax}_{1 \leq i \leq N} \sum_{j=1, j \neq i}^N w_{i,j} \quad (4)$$

Based on the cosine similarity values between the core sentence and all other sentences, all sentences in the document are divided into a subject relevant sentence set and an irrelevant sentence set. The subject relevance score of each word is based on the frequency of appearance in the subject relevant and irrelevant sentence sets. The word relevance score RS_d is calculated by Equation (5) (Jeong et al., 2016).

$$\begin{aligned} RS_d(word) &= \log \frac{p_d \times (1 - q_d)}{(1 - p_d) \times q_d} \\ &= \log \frac{(r_d + 0.5)(S_d - s_d + 0.5)}{(R_d - r_d + 0.5)(s_d + 0.5)} \end{aligned} \quad (5)$$

where p_d and q_d are the probabilities that a word appears in subject relevant and irrelevant sentences set in a document d . Respectively, R_d and S_d are the number of subject relevant sentences and irrelevant sentences in document d . r_d and s_d are the number of subject relevant and irrelevant sentences that include the word in a document d , and 0.5 is a naïve smoothing factor to avoid zero-denominator or log-zero. On this paper, the top 30% of sentences with the subject relevance score were selected as the subject relevant sentence set, and the others were as the irrelevant sentence set. Afterward, the sentence score can be calculated with the sum of relevance score of words included in the sentence (Jeong et al., 2016).

$$Score(s_i) = \sum_{word \in s_i} RS_d(word) \quad (6)$$

3.3 Fake News Detection

After we calculate the sentence ranking using the sentence scores by Equation (6), we can construct a new updated graph by reflecting the sentence ranking into the weight of the edge. The weight of the edge represents the reflection rate of the subject

information as well as the context information.

$$R_{Score(s_i)} = 1 - \frac{Sentence_Rank(s_i) - 1}{N} \quad (7)$$

$$w'_{i,j} = w_{i,j} * R_{Score(s_i)} \quad (8)$$

(for $i, j = 1, 2, \dots, N$ and $i \neq j$)

The sentence ranking score is calculated by Equation (7) and the weight of all edges is updated by Equation (8). Subsequently, we can construct subject contextualized sentence embeddings using the updated graph. The subject contextualized sentence embedding (i.e., h'_i) is created by the weighted sum of the initial sentence embedding of the current sentence (i.e., h_i) and those of the other adjacent sentences (i.e., h_j) as follows:

$$h'_i = h_i + \sum_{j=1, j \neq i}^N w'_{i,j} * h_j \quad (9)$$

Finally, we create a document embedding by averaging the embedding vectors of all sentences in a document and then the document embedding is then fed into a multi-layer feedforward neural network to predict label, fake or real, by the binary classification, as a vector \hat{y} .

$$doc = \sum_{i=1}^N h'_i \quad (10)$$

$$\hat{y} = \operatorname{Softmax}(\tanh(Wdoc + bias)) \quad (11)$$

The cross-entropy loss function is used to optimize our neural network.

4 Experiments

4.1 Datasets

We used HDSF dataset (Karimi and Tang, 2019) for our fake news detection experiments. The dataset consists of 3,360 real documents and 3,360 fake documents. We follow the Karimi's data split: 6,452 documents for training data, 134 ones for validation set, and 134 ones for test set and each data split contains even number of documents from fake and real classes. In addition, we did experiments on 5-folds cross-validation because the original HDSF dataset contain too small size of test documents.

4.2 Experimental Settings

We used word2vec embedding (Mikolov et al., 2013) that is pre-trained from Google as an initial word embedding, and set the Bi-LSTM hidden unit

size to 200. We set the epoch as 200, mini-batch size as 40, and dropout as 30% on each experiment. We used the Adam optimizer (Kingma and Ba, 2014) and set the initial learning rate as 0.001 and reduce by 10 times for every 50 epochs. We used accuracy as the metric of performance.

4.3 Models

The proposed model is compared to the baseline model and Karimi’s model (Karimi and Tang, 2019) to prove a superiority of our proposed model in this subsection.

Karimi’s Model (Karimi and Tang, 2019) A fake detection method to predicts whether a document is fake or real by constructing relationships of each sentence in a document using the hierarchical discourse-level dependency tree.

Baseline Model The baseline model uses the document embedding only using Bi-LSTM and predicts whether it is fake or real through the feed-forward neural network.

Graph Model This graph model connects all sentences in a document by constructing a graph structure. Fake news is detected by creating sentence embeddings that only reflect contextual information on a graph.

Graph + Summarization Model (Proposed) The graph + summarization model is our final proposed model. It classifies the fake or real documents after representing a document embedding that reflect the contextual and subject information on a graph structure by the summarization technique.

4.4 Experiment Results

Table 1 shows a comparison of the experiment results. The RST, LIWC, N-grams, and BiGRNN-CNN models implemented by (Karimi and Tang, 2019). Our proposed model showed best performance compared to other models and it achieved better performance than the baseline model about 11.19%p, and even better than the Karimi’s model about 8.85%p. In addition, we obtained 3.73%p improvement when summarization technique is used in the graph model. Moreover, the final proposed model showed the best performance even in the 5-folds cross-validation experiments with 9.06%p improvement as well.

5 Analysis

Herein, we introduce two analyses by an additional experiment for subject consistency detection in our

Model	Data	
	HDSF	Cross Validation
RST	67.68	-
LIWC	70.26	-
N-grams	72.37	-
BiGRNN-CNN	77.06	-
Karimi’s	82.19	81.71
Baseline	79.85	79.27
Graph Model	87.31	84.52
Graph	91.04	88.33
+ Summarization (Proposed)	(+11.19%p*)	(+9.06%p*)
	(+8.85%p*)	(+6.62%p*)

Table 1: Comparing results. The * and * denote the differences between the proposed model and the Baseline or Karimi’s.

model and dataset. The proposed method attempted to effectively detect subject information in fake news and the consistency of subject information is also important to detect fake news. In the first analysis, we verify our process of updating the graph in the proposed model, which uses a fully connected graph, by comparing other model using a 50% connected graph, which uses edges with only the top 50% with high attention scores. As a result, the proposed model with a fully connect graph showed 1.29%p higher performance than the model with a 50% connected graph. We think it means that a fully connect graph is more useful to detect fake news. Secondly, we observed the variances of the attention scores in the fake and real documents (see Table 2). The variance in the fake documents is higher than that of the real documents and it means that the fake documents has an inconsistent subject distribution.

	Variance	Standard Deviation
Fake	8.6	2.93
Real	5.41	2.33

Table 2: Variance/standard deviation of attention scores

6 Conclusions

We have proposed a novel graph based fake news detection method using summarization technique. Our model shows that the use of contextual and subject information is helpful in detecting fake news. Our final proposed model achieved better performance than the baseline model about 11.19%p.

Acknowledgements

This work was supported in part by Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (MSIT) (No. 2020-0-00368, A Neural-Symbolic Model for Knowledge Acquisition and Inference Techniques), in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C2100362), and in part by the MSIT (Ministry of Science and ICT), Korea, under the ICT Creative Consilience program (IITP-2020-0-01821) supervised by the IITP.

References

- Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684.
- Timothy Dozat and Christopher D. Manning. 2018. [Simpler but more accurate semantic dependency parsing](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 484–490, Melbourne, Australia. Association for Computational Linguistics.
- Hyoungil Jeong, Youngjoong Ko, and Jungyun Seo. 2016. How to improve text summarization and classification by mutual cooperation on an integrated framework. *Expert Systems with Applications*, 60:222–233.
- Hamid Karimi and Jiliang Tang. 2019. [Learning hierarchical discourse-level structure for fake news detection](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3432–3442, Minneapolis, Minnesota. Association for Computational Linguistics.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Or Levi, Pedram Hosseini, Mona Diab, and David Broniatowski. 2019. [Identifying nuances in fake news vs. satire: Using semantic and linguistic cues](#). In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 31–35, Hong Kong, China. Association for Computational Linguistics.
- Yi-Ju Lu and Cheng-Te Li. 2020. [GCAN: Graph-aware co-attention networks for explainable fake news detection on social media](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 505–514, Online. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. 2019. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*.
- Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. [A stylistic inquiry into hyperpartisan and fake news](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 231–240, Melbourne, Australia. Association for Computational Linguistics.
- Rui Yan, Ian EH Yen, Cheng-Te Li, Shiqi Zhao, and Xiaohua Hu. 2015. Tackling the achilles heel of social networks: Influence propagation based language model smoothing. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1318–1328.
- Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. 2012. Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD workshop on mining data semantics*, pages 1–7.