2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application

Perceiving the Narrative Style for Fake News Detection Using Deep Learning

Mengtao Sun
dept. ICT and Natural Sciences
Norwegian University of Science and Technology
Ålesund, Norway
mengtao.sun@ntnu.no

Ibrahim A. Hameed
dept. ICT and Natural Sciences
Norwegian University of Science and Technology
Ålesund, Norway
ibib@ntnu.no

Hao Wang
dept.Computer Science
Norwegian University of Science and Technology
Gjøvik, Norway
hawa@ntnu.no
* Corresponding Author

Mark Pasquine
dept. International Business
Norwegian University of Science and Technology
Ålesund, Norway
mapa@ntnu.no

Abstract— **Existing deep-learning-based features have shown strong enough results (more than 90% accuracy) if a large amount of annotated data is available. However, in reality, data annotation is labor-intensive and expensive. This work proposes a novel approach for fake news detection by perceiving narrative style with deep learning to alleviate the problem. Deep-learning-based features are represented as embedding vectors. Traditional embedding vectors ingest word context information, but the training requires considerable annotated datasets. Many linguistics studies have shown that written styles, such as the usage of punctuation, repetition of words, and grammatical order, are significant for distinguishing fake news. Our model takes advantage of the word-to-word dependency relationship, describing the styles of the news utterances. We denote the proposed model as Syntax Graphical Thread (SGT) network. We utilize a trainable randomly initialized embedding and Gated Recurrent Unit (GRU) layer to capture the context vector, while a Graph Attention (GAT) layer is used to capture the narrative features. The experimental result manifests that our method can significantly mitigate the reliance on data scale and present better classification results when the dataset is limited.**

Keywords—Fake News Detection, Gate Recurrent Unit, Graph Attention Network, Dependency Parsing

I. Introduction

Fake news detection is a typical binary classification problem. It has shown strong enough results (over 90%) when the data scale is sufficient [1]. In recent years, there have been many studies on the improvement of deep neural networks such as Convolutional Neural Network (CNN) [2], Gated Recurrent Unit (GRU), and ensemble methods [3, 4]. However, on the other hand, these methods rely heavily on the data scale. If there is less labeled data, the expressive ability of AI-driven methods will be hindered. For a dataset with limited annotations, machine learning methods are still ahead of deep learning models in some cases [5]. Therefore, this research mainly focuses on improving fake news detection with limited annotation by neural networks.

In deep learning, news features are based on context information. That is, the model input is word embedding [6, 7]. However, word embedding cannot generalize content well and produce a suitable embedding vector when the data scale is low. To obtain more discriminative information between true and fake news. Some linguistical factors are considered [8], such as punctuation and sentence length. Statistical measurements can poorly be integrated with the trainable idea of deep learning. We suggest that there are differences in the narrative style of true and fake news. We hope to add style embedding into the definitive word embedding to improve the distinction of news content, especially when the data scale is not enough.

When the text classifications are carried out, people usually ignore the prior knowledge of syntax grammar. With the development of graph neural networks in recent years, the graphical data frame can be better analyzed. On the one hand, the text belongs to the sequential data frame, which is from front to back. On the other hand, grammatical parsing can also be expressed as a tree data frame, and the graphical data frame comprises the information about word dependency, word position, and word order. Fig. 1 exemplifies the graphic text frame of a news utterance.

In Fig. 1, 'approves' is the root of the utterance. 'Thailand', 'billion' are the child nodes of the root 'approval'. '2.2', '$' and

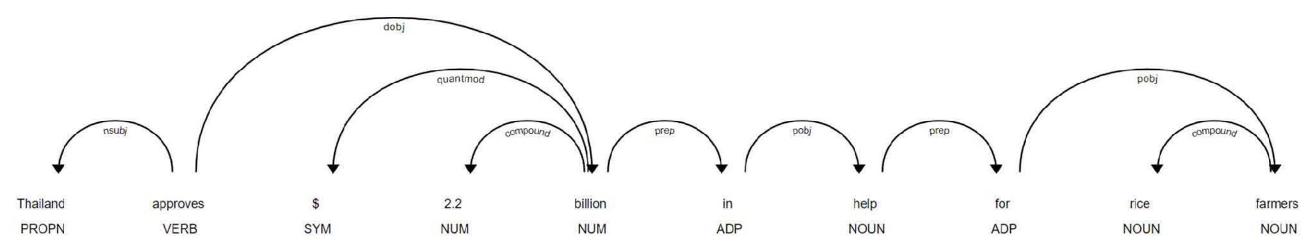

Fig. 1. An Example of Syntax Dependency Parsing Result

978-1-6654-9457-1/21/$31.00 ©2021 IEEE
DOI 10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00184

2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys) | 978-1-6654-9457-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/HPCC-DSS-SMARTCITY-DEPENDSYS53884.2021.00184

other descendant nodes of 'approval' are connected with each other with the graphical data frame according to the dependency. In this way, the graph structure can generalize the style information from real and fake news and strengthen the degree of discrimination.

This paper comprehensively considers the context and the narrative style of news text. The main contributions of this paper are as follows:

1) This paper combines context embedding and narrative embedding to detect fake news.

2) We propose a novel neural network incorporating the abovementioned **S**yntax **G**raphical **T**hread, call SGT network.

3) We compare the performance with different baselines used in low-resource text classification. We also modify the news length and data scales during the comparison.

The experimental results show that the SGT network effectively integrates with narrative styles, so that improves the performance of fake news detection. We then analyze the role of each component in SGT network by exploratory experiments.

This article is distributed as follows: Section 2 introduces the related research of fake news detection. Section 3 describes the overall flow and details of the proposed SGT network. Section 4 illustrates the experiments and the results.

II. Related Work

Various researchers have explored and studied methods for fake news detection by using the content or social context information [7]. Here, SGT Network focuses on content-based features. Content-based features are primarily extracted from the news text. In addition to the usual feelings or emotions [9], it can also be observed from the text characteristics that a clear written style appears in fake news content [11]. Choudhary proposed a linguistic-based fake news detection model including syntactic, sentimental, grammatical, and readability features [8].

Content-based models of fake news detection include machine learning and deep learning. Ahmad et al. compared the performance of various statistical machine learning models in fake news classification experiments [12]. Cuşmaliuc et al. use support vector machines, naive Bayes, and random forest methods to analyze the inner differences between these methods and apply them for twitter fake news detection [13]. It is undeniable that statistical machine learning methods still play an essential role. More research is focused on the deep learning model, especially the improvement based on CNN and GRU. Goldani et al. [14] improved the fake news recognition performance of a Convolutional neural network by designing a new margin loss. Nasir et al. [15] proposed a hybrid algorithm based on CNN and GRU and achieved higher evaluations. To be more complex. Kumar et al. proposed a hybrid model combining an Attention Bi-GRU with CapsNet. The model produced better results for cyberbullying detection on social media [16]. Dun et al. proposed a novel knowledge-aware attention network that incorporated external knowledge from knowledge graphs for fake news detection [17]. Zhang et al. proposed a graph neural network to address the unknown characteristics of fake news using diverse connections among news articles, creators, subjects [18]. Sun et al. proposed a merging technology that

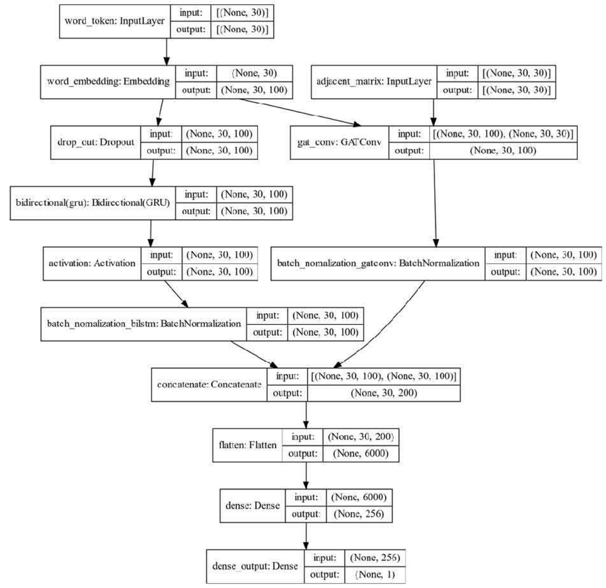

Fig. 2. Structure of SGT Network

extracts the content features from two different embeddings for text classification [19].

For social context-based methods, these features include (i) user-based features, (ii) post-based features, and (iii) web-based features. User-based features are extracted from user profiles to measure the potential fake attributes [20]. The post-based features highlight the user's social participation in various positions [21] and credibility [22]. Network-based features are mainly extracted by constructing accurate detection systems, such as diffusion networks and correlation networks. [23]. With the widespread adoption of social media, fake news detection research also considers social media activities for detecting fake news, for example, early detection through social learning [24] and user relationship.

III. Method

In this Section, we introduce our network structure and each component in detail. We use a concise method to acquire the knowledge of context and narrative style. We use the TensorFlow platform to achieve our idea. The structure generated by TensorFlow is visually displayed in Fig. 2.

A. Overall Structure

SGT network is designed in parallel embedding mode. From Fig. 2, we can observe the computational flow of SGT model. It has been confirmed that GRU is particularly effective for text data [25]. In tradition, fake news detection has been investigated by adopting a standard text classification model that consists of an Embedding layer as input in the form of word embedding vectors, followed by a bi-directional GRU (bi-GRU) layer, and a predictive dense layer. The design of our model is motivated by the concept of multiple parallel channels-variable-size-based neural networks [26]. Our proposed model reaps the benefits of both traditional context features and written style by syntax dependency parsing [27]. In our proposed model, inputs comprise an extra adjacent matrix obtained from syntax dependency parsing. We first use randomly initialized embedding to discover the features of context, sending word embedding vectors to the dropout layer and then to the bidirectional GRU layer. In the other parallel route (see gat_conv in Fig 2.), we send the word embedding vector and the adjacent matrix to the graph attention network used to produce the narrative style embedding.

B. Model Components

Next, we discuss the SGT components, viz., the importance of each deep neural layer and why we use these components in neural networks. These layers include Embedding, bi-GRU, GAT, dropout, batch normalization, and activation function selection.

Embedding Embeddings generally represent geometrical encodings [28] of words based on how frequently they appear together in a text corpus. In our paper, embeddings comprise the context information from given news. To verify the performance of using syntaxial style between fake news and true news, we do not use the pretrained embeddings.

Dropout Dropout is a regularization technique [29], which aims to reduce the complexity of any model with the end goal of preventing over-fitting. After the embedding layer, we applied a dropout layer because the Embedding is easy overfitting in the limited dataset. By experiments, the dropout value is 0.5 throughout our experiments, i.e., given a layer that makes 50% value in embedding vector to zero in training.

Bi-GRU Bi-GRU layer is widely applied to serialize sentences in forward and backward order. Bi-GRU contains two directional information and achieves a more satisfactory performance compared with unidirectional GRU. The outputs of the Bi-GRU layer represent the knowledge of context. In the current era of computing, GRU as the classical sequential model shows better performance than Simple Recurrent Neural Network (RNN) and trains faster than Long-Short-Term-Memory Network (LSTM).

Batch Normalization Although we force the outputs of both bi-GRU and GAT layer with identical dimensions, we suppose their outputs are not equally stable. A good way to address the problem is to re-center and re-scale the values according to normalization. We equip a Batch Normalization function followed by both bi-GRU and GAT layer. The performances with Batch Normalization and without are compared in Section 4.

Adjacent Matrix The adjacency matrix is a square matrix used to represent the parsing tree from the news sentence. We use the SpaCy library to obtain the parsing results automatically [30]. The English dependency parsing was pretrained from OntoNotes 5, WordNet 3.0, and GloVe Common Crawl [30]. To investigate the best representation of word-to-word dependency, we separately perform directional and bi-directional dependency relationships. The results in Section 4 show that bi-directional dependency parsing can more effectively capture the written features.

Graph Attention Network Graph neural network is successfully applied to node classification tasks [31, 32]. GAT is a type of Graph Neural Network (GNN) that shows the best performance in recent applications. We use the same GAT structure as [31]. Given the graphical node-to-node relationship, GAT will generate the outputs, which contain sentence dependency grammars and features from preceding layers.

Concatenation & Flatten The concatenation layer combine the GRU output and GAT output according to their feature channels. The flatten layer converts the features taken from the concatenated vector and maps it to a single column, further passed to the fully connected layer.

Dense Layer / Fully Connected Layer We can understand the functionality of a dense layer as a linear operation in which every input is connected to every output by trainable weights. We employ two dense layers following the flatten layer. The first dense layer aims at mixing the combined features from context and writing style. The second dense layer perceive the probability of true or fake news. Results close to 1 represent true news while 0 represent fake news.

Activation The commonly used activation function is rectified linear unit (ReLU) [33]. The main functionality of ReLU is that it successfully removes negative values from an activation map by setting them to zero in a network. ReLU also increases the nonlinear properties of the decision-making function in the complete network. In this study, we use exponential linear units (ELU) [34] as activation function in each layer. With the same functionality of ReLU, an ELU activation function tends to converge errors to zero faster and produce more accurate results in real tasks compared with RELU [34].

We use Adam optimizer. The learning rate is 0.0001 for training with cross entropy loss. The batch size is 64, and we set training epochs as 100. The vocabulary list contains 5000 words. The sentence length ranges from 30 to 80 tokens..

C. Model Metrics

To evaluate the performance of our proposed model, we applied standard Accuracy (Acc), Precision (P), Recall (R), F1 Score between the model output and ground truth. These metrics are calculated based on the binary Confusion Matrix. The definitions of factors are given in Table I.

TABLE I. DEFINITION OF CONFUSION MATRIX

	Predicted Positive	***Predicted Negative***
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

According to confusion matrix, the four metrics are defined as:

Recall $$R = \frac{TP}{TP+FN} \tag{1}$$

Precision $$P = \frac{TP}{TP+FP} \tag{2}$$

F1 Score $$F1 = \frac{2*(P*R)}{(P+R)} \tag{3}$$

Accuracy $$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

IV. EXPERIMENT

In this section, we describe the datasets for the experiment and briefly introduce the baselines. The experiments are distributed in the following order:

We demonstrate the main results of SGT in a relatively low resource dataset (400 instances with 30 utterance length). The training, development and testing ratio is 0.6, 0.2, 0.2. The model achieving best results on the development set was chosen for the final evaluation on the test set. We then probe the influences of SGT components on the development set. Next, we gradually increase the utterance length and data scale to compare the effectiveness of SGT on the limited dataset.

A. Dataset

Experiments were conducted using the ISOT fake news dataset[1] [35, 36] . The dataset contains two types of articles fake and real News. This dataset was collected by University of Victoria. The truthful articles were obtained by crawling articles from Reuters.com. The fake news articles were collected from unreliable websites that were flagged by Politifact (a fact-checking organization in the USA) and Wikipedia. The majority of articles focus on politics.

The description of the ISOT dataset is shown in Table II. True news contains 21,417 articles. Fake news contains 23,481 articles. The collected articles are ranged from 2016 to 2017. We randomly extract 400 articles a time and repeat each experiment 5 times to get the average performance.

TABLE II. DATASET DESCRIPTION

News	*Size*	*Subjects*	*Size*
Real-News	21,417	World News	10,145
		Politics News	11,272
Fake-News	23,481	Government News	1,570
		Middle-east	778
		US News	783
		left-news	4,459
		Politics	6,841
		Others	9,050

TABLE III. PARAMETERS OF BASELINES

CNN	Value	Bi-GRU	Value
Embedding size	100	Embedding size	100
Dropout Rate	0.5	Dropout Rate	0.5
No. CNN layer	2	No. bi-GRU layer	2
No. Dense layer	2	No. Dense layer	2
Optimizer	Adam	Optimizer	Adam
Activation	ELU	Activation	ELU
Learning Rate	0.001	Learning Rate	0.001
Epoch	5	Epoch	5

[1] https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php

TABLE IV. PERFORMANCES OF SGT AND BASELINES

Model	*F1*	*R*	*P*	*Acc*
Random Forest	82.49	82.49	82.49	82.50
Logistic Regression	79.68	79.74	81.30	80.00
Decision Tree	76.22	76.20	76.25	76.25
Gaussian NB	69.78	71.79	82.54	72.50
Linear Discriminant Analysis	76.30	76.99	82.91	77.50
Support Vector Machine	78.34	78.46	80.33	78.75
Multinomial Naive Bayes	80.36	80.77	**86.61**	81.25
Convolutional Unit	76.25	77.62	77.81	76.25
Bidirectional Gated Recurrent Unit	73.41	76.35	79.56	73.75
SGT	**82.50**	**82.97**	82.97	**82.50**

TABLE V. ABLATION STUDY

SGT	*F1*	*R*	*P*	*Acc*
origin	**92.5**	**92.4**	**92.7**	**92.5**
- Symmetry Adjacent Matrix	74.5	74.1	75.1	76.3
- Bi-direction	82.1	81.8	82.8	82.5
- BatchNorm(Bi-GRU)	82.1	81.9	82.4	82.5
- BatchNorm(GAT)	84.9	85.0	85.7	85.0
- GRU	77.8	77.5	81.7	77.5

B. Baseline Models

Machine Learning: We conduct 7 typical machine learning methods as baselines which are effectively used in a moderate fake news detection dataset.

Deep Learning: The most effective components of deep learning are convolutional units and gated recurrent units. We separately apply the convolution and gated recurrent unit with dense layers as baselines. The parameters of baselines are shown in Table III.

C. Main Results

It can be seen from Table IV that in the case of limited news instances, the machine learning model is overall ahead of the deep learning model. Among them, the Multinomial Naive Bayes has the best Precision, and the Random Forest got the best Recall, Accuracy and F1 score. In the case of limited dataset, the performance of the deep learning models using the convolutional unit and the gated recurrent unit is equivalent to the average performance of the machine learning model. In conclusion, relying solely on context-based features cannot well fit the parameters of the deep learning model, which requires us to increase the inputted features. By applying the narrative style we defined, the results are improved overall. The SGT model can be simply regarded as a fusion model of traditional bi-directional GRU + Narrative style embedding. Compared with no narrative style embedding, i.e., the sole bi-GRU model, SGT improves 9.09 on F1 score, 6.62 on Recall, 3.41 on Precision and 8.75 on Accuracy. The performance is also generally better than all machine learning models.

D. Ablation Study

In order to study the contribution of each component in SGT, we conducted ablation experiments on the data set, and the results are shown in Table V. By default, we use a symmetric adjacency matrix, that is, if word A and B have a dependency relationship, then B and A have a dependency relationship. If we use one-direction dependencies, we find that the performance of the model is severely degraded. It shows that bi-directional dependence plays an important role in the performance of the model. We also found that if the bi-directional GRU is changed to GRU, the four evaluations of the model are reduced by about 10%, indicating that the bi-direction is important in the language model. In addition, without the help of Batch Normalization on the intermediate output of Bi-GRU and GAT, the result will have a great impact. If we directly remove the bi-GRU layer in the SGT network, the model is simplified to a network that relies solely on GAT. The results show that the expressiveness of the model is greatly reduced, indicating that the sequential model provides more features for fake news detection. As a result, by combining context and narrative style, SGT has stronger sentence modeling and feature acquisition capabilities for discrimination of true and fake news.

E. Performance against Sentence Length

Fig. 3-6 shows the performance of SGT and several baseline models on different news sentence lengths. All the experiments are trained in 400 randomly selected news instances. We divide the data set into 11 parts based on sentence length. The sentence length ranges from 30 to 80. The dotted line model is a model based on machine learning, the solid line model is a model based on deep learning, and the solid line with the small circle is our proposed method.

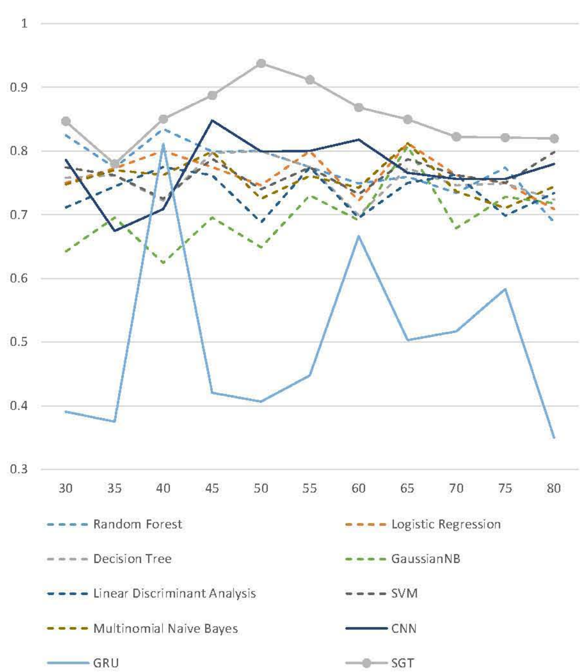

Fig 3. F1 score: Performance against Sentence Length

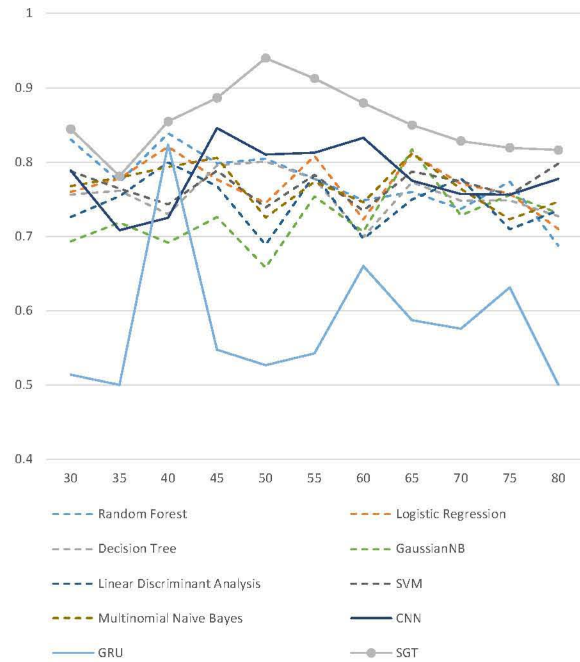

Fig 4. Recall (R) : Performance against Sentence Length

The statistical machine learning models are more stable at different sentence lengths, but the performances always stay at a medium level. The CNN model performs best when the

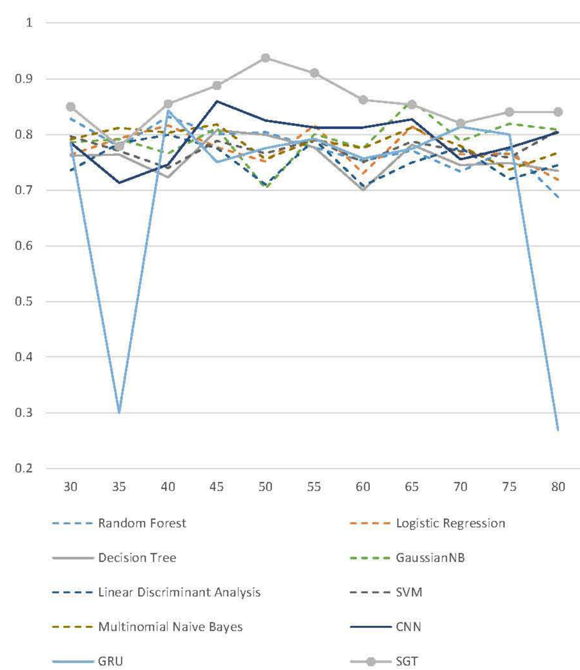

Fig 5. Precision (P) : Performance against Sentence Length

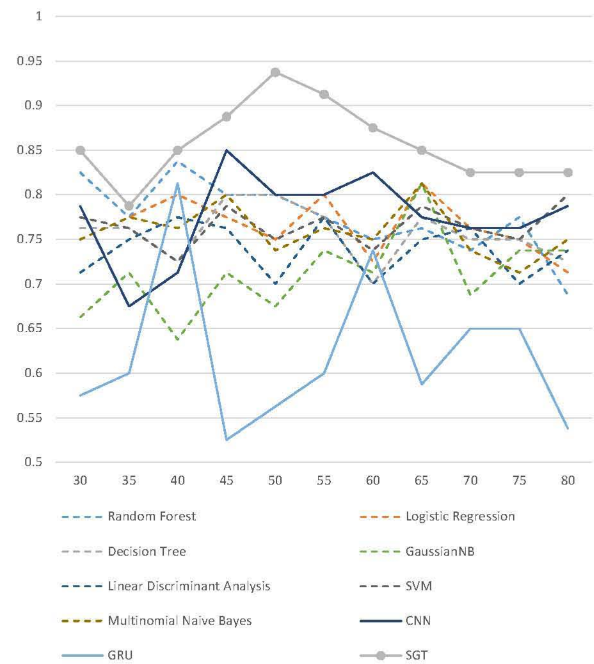

Fig 6. Accuracy (Acc) : Performance against Sentence Length

sentence lengths are ranged from 45 to 60, but they become worse when the sentence lengths are below or above the scope. The bi-GRU model has the worst stability, and the overall performance is the worst among others. SGT not only gives higher results on short sentences but also shows its effectiveness and robustness when the sentence length exceeds 60 tokens. Compared with the baseline, it gives higher Precision, Recall, Accuracy and F1 scores, which shows that the SGT model can better capture sentence features between fake and true news.

F. Performance against the number of News Instance.

Fig. 7-10 shows the performance of SGT and several baseline models on different data scales. All experiments are based on 30 sentence-length news examples. The data scale ranges from 200 to 1500 instances, and each experiment is conducted by gradually adding 100 news instances. Random forest is a powerful baseline in machine learning models, which is superior to other machine learning models at different scales. However, with the length increasing, the accuracy of the random forest decreases in four metrics. The evaluation of other machine learning models has also declined when the data size exceeds 1,000, neither F1 score nor Accuracy exceeds 80%. In contrast to machine learning models, deep learning models have increased ability on feature capture when the data is sufficient. The bi-GRU and CNN-based methods can finally reach a performance of 90% in the four evaluation indicators. It is worth noting that regardless of the scales of the experimental data, SGT has shown satisfactory results. SGT leads other baselines in small-scale data and has no drop in large-scale data.

V. CONCLUSION

In this work, we propose a narrative-style-based method to expand the content feature of true and fake news to address the

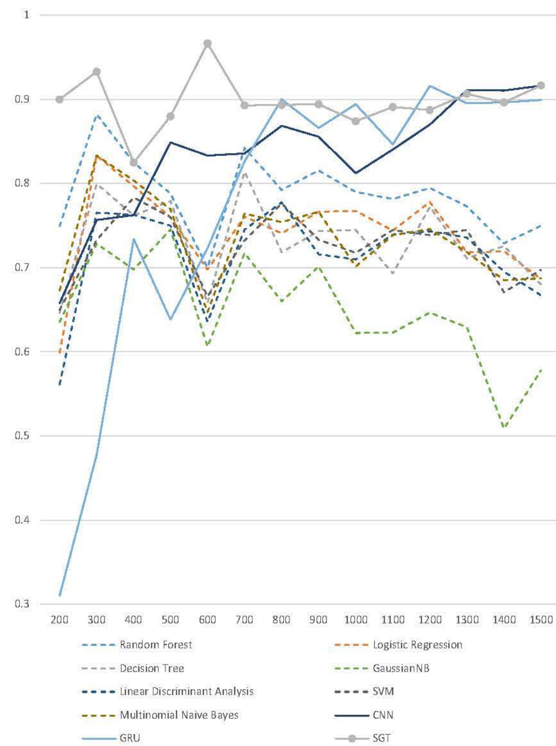

Fig 7. F1 score: Performance against the Number of News Instance

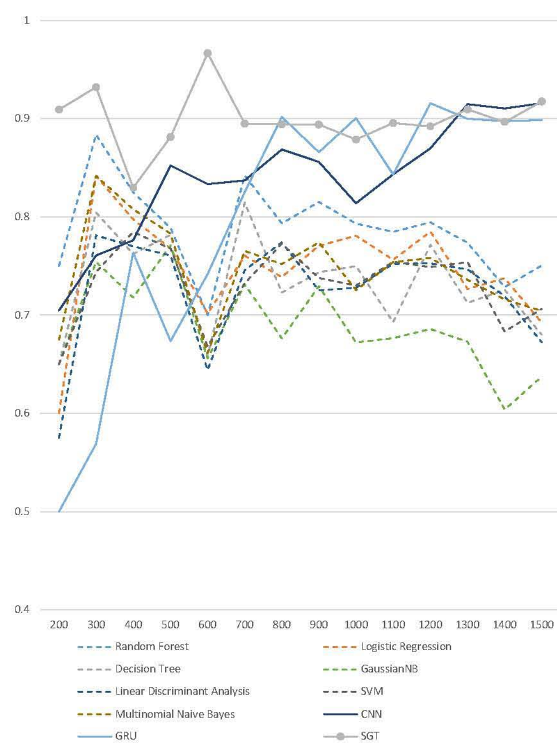

Fig 8. Recall (R) : Performance against the Number of News Instance

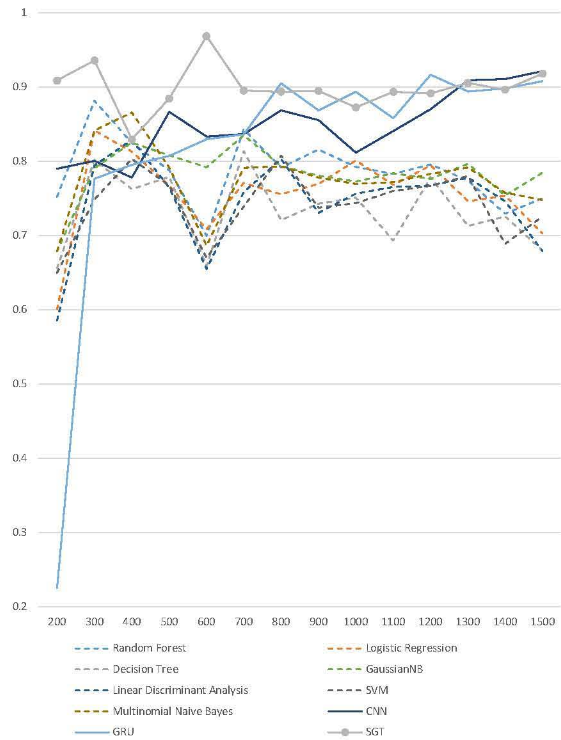

Fig 9. Precision (P) : Performance against the Number of News Instance

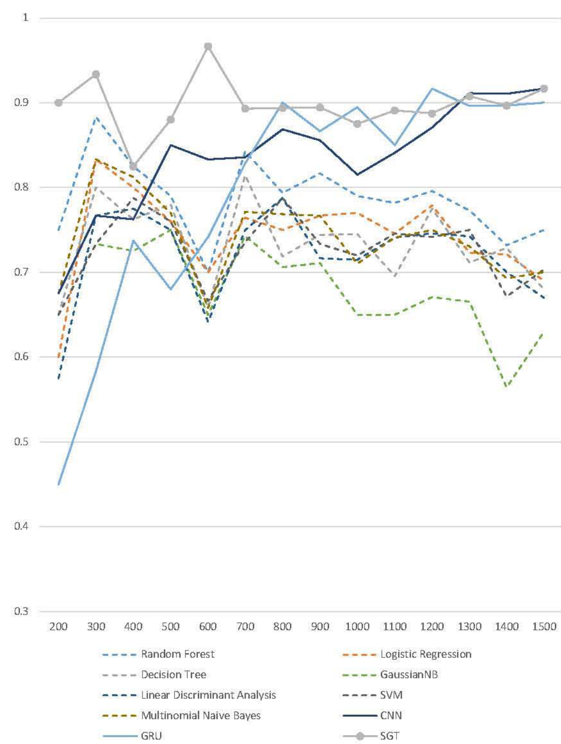

Fig 10. Accuracy (Acc) : Performance against the Number of News Instance

poor performance of news classification under low resources. We use syntactic dependency to obtain the written style of the news article and add these features with word-level context features. SGT supports fake news detection with different sentence lengths, and it is less dependent on the size of the data. It shows the best performance in comparison experiments with other models. The explanatory experiment also illustrates the influence of different components in the SGT model on the output results.

In the future, we will improve the fusion method to better integrate the narrative style with word-level context features. We will also extend the application of narrative style vectors and improve the performance of content-based text classification tasks.

REFERENCES

[1] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi. "Evaluating deep learning approaches for Covid19 fake news detection." In International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation, pp. 153-163. Springer, Cham, 2021.

[2] Q. Li, Q. Hu, Y. Lu, Y. Yang, and J. Cheng. "Multi-level word features based on CNN for fake news detection in cultural communication." Personal and Ubiquitous Computing 24, no. 2 (2020): 259-272.

[3] N. Aslam, I. U. Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil. "Fake Detect: A Deep Learning Ensemble Model for Fake News Detection." Complexity 2021 (2021).

[4] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. KR. Maddikunta, and W. Z. Khan. "An ensemble machine learning approach through effective feature extraction to classify fake news." Future Generation Computer Systems 117 (2021): 47-58.

[5] F. I. Adiba, T. Islam, M. S. Kaiser, M. Mahmud, and M. A. Rahman. "Effect of corpora on classification of fake news using naive Bayes classifier." International Journal of Automation, Artificial Intelligence and Machine Learning 1, no. 1 (2020): 80-92.

[6] N. A. Smith. "Contextual word representations: putting words into computers." Communications of the ACM 63, no. 6 (2020): 66-74.

[7] J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu. "Content based fake news detection using knowledge graphs." In International semantic web conference, pp. 669-683. Springer, Cham, 2018.

[8] A. Choudhary, and A. Arora. "Linguistic feature based learning model for fake news detection and classification." Expert Systems with Applications 169 (2021): 114171.

[9] Y. Liu, and Y. B. Wu. "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks." In Thirty-second AAAI conference on artificial intelligence. 2018.

[10] S. Kula, M. Choraś, R. Kozik, P. Ksieniewicz, and M. Woźniak. "Sentiment analysis for fake news detection by means of neural networks." In International Conference on Computational Science, pp. 653-666. Springer, Cham, 2020.

[11] S. Ghosh, and C. Shah. "Towards automatic fake news classification." Proceedings of the Association for Information Science and Technology 55, no. 1 (2018): 805-807.

[12] I. Ahmad, M. Yousaf, S. Yousaf, and M. O Ahmad. "Fake news detection using machine learning ensemble methods." Complexity 2020 (2020).

[13] C. Cuşmaliuc, L. Coca, and A. Iftene. "Identifying Fake News on Twitter using Naive Bayes, SVM and Random Forest Distributed Algorithms." In Proceedings of The 13th Edition of the International Conference on Linguistic Resources and Tools for Processing Romanian Language (ConsILR-2018) pp, 177-188. 2018.

[14] M. H. Goldani, M. Hadi, R. Safabakhsh, and S. Momtazi. "Convolutional neural network with margin loss for fake news detection." Information Processing & Management 58, no. 1 (2021): 102418.

[15] J. A. Nasir, O. S. Khan, and I. Varlamis. “Fake news detection: A hybrid CNN-RNN based deep learning approach.” International Journal of Information Management Data Insights 1, no. 1 (2021): 100007.

[16] A. Kumar, and N. Sachdeva. “A Bi-GRU with attention and CapsNet hybrid model for cyberbullying detection on social media.” World Wide Web (2021): 1-14.

[17] Y. Dun, K. Tu, C. Chen, C. Hou, and X. Yuan. “KAN: Knowledge-aware Attention Network for Fake News Detection.” In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 1, pp. 81-89. 2021.

[18] J. Zhang, B. Dong, and S. Y. Philip. “Fakedetector: Effective fake news detection with deep diffusive neural network.” In 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 1826-1829. IEEE, 2020.

[19] M. Sun, I. A. Hameed, and H. Wang. “A Novel Ensemble Representation Framework for Sentiment Classification.” In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2020.

[20] R. Basak, S. Sural, N. Ganguly, and S. K. Ghosh. “Online public shaming on Twitter: Detection, analysis, and mitigation.” IEEE Transactions on Computational Social Systems 6, no. 2 (2019): 208-220.

[21] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein. “A stylometric inquiry into hyperpartisan and fake news.” Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics Volume 1: Long Papers. 2018.

[22] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu. “Unsupervised fake news detection on social media: A generative approach.” In Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 01, pp. 5644-5651. 2019.

[23] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang. “Multi-source multi-class fake news detection.” In Proceedings of the 27th international conference on computational linguistics, pp. 1546-1557. 2018.

[24] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias. “Enhancing deep learning sentiment analysis with ensemble techniques in social applications.” Expert Systems with Applications 77 (2017): 236-246.

[25] Y. Liu, A. Pei, F. Wang, Y. Yang, X. Zhang, H. Wang, H. Dai, L. Qi, and R. Ma. “An attention-based category-aware GRU model for the next POI recommendation.” International Journal of Intelligent Systems (2021).

[26] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha. “FNDNet–a deep convolutional neural network for fake news detection.” Cognitive Systems Research 61 (2020): 32-44.

[27] Y. Qiao, D. Wiechmann, and E. Kerz. “A language-based approach to fake news detection through interpretable features and brnn.” In Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media (RDSM), pp. 14-31. 2020.

[28] X. Zhang, J. Zhao, and Y. LeCun. “Character-level convolutional networks for text classification.” Advances in neural information processing systems 28 (2015): 649-657.

[29] B. Zhong, X. Xing, P. Love, X. Wang, and H. Luo. “Convolutional neural network: Deep learning-based classification of building quality problems.” Advanced Engineering Informatics 40 (2019): 46-57.

[30] M. Honnibal, I. Montani, S. V. Landeghem, and A. Boyd. “spaCy: Industrial-strength Natural Language Processing in Python (2020).” URL https://doi. org/10.5281/zenodo 1212303 (2020).

[31] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. “Graph attention networks.” International Conference on Learning Representations (ICLR 2018).

[32] Q. Huang, J. Yu, J. Wu, and B. Wang. “Heterogeneous graph attention networks for early detection of rumors on twitter.” In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2020.

[33] Y. Li, and Y. Yuan. “Convergence analysis of two-layer neural networks with relu activation.” Advances in neural information processing systems (2017), pp. 597-607

[34] D. Clevert, T. Unterthiner, and S. Hochreiter. “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”. In Proceedings of the International Conference on Learning Representations. 2016.

[35] H. Ahmed, I. Traore, and S. Saad. “Detecting opinion spams and fake news using text classification.” Security and Privacy 1, no. 1 (2018): e9.

[36] H. Ahmed, I. Traore, and S. Saad. “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, 2017.