Fake News Detection in Social Media based on Multi-Modal Multi-Task Learning

Xinyu Cui
Northeast Forestry University
Harbin, China

Yang Li*
Northeast Forestry University
Harbin, China

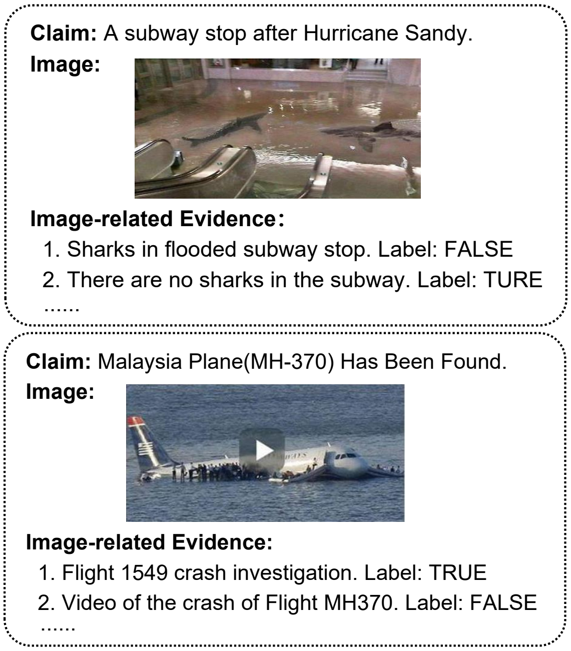
Abstract—**The popularity of social media has led to a substantial increase of data. The task of fake news detection is very important, because the authenticity of posts cannot be guaranteed. In recent years, fake news detection combining multi-modal information such as images and videos has attracted wide attention from scholars. However, the majority of research work only focuses on the fusion of multi-modal information, while neglecting the role of external evidences. To address this challenge, this paper proposes a fake news detection method based on multi-modal and multi-task learning. When learning the representation of the news posts, this paper models the interaction between images and texts in posts and external evidences through a multi-level attention mechanism, and uses evidence veracity classification as an auxiliary task, so as to improve the task of fake news detection. Authors conduct comprehensive experiments on a public dataset, and demonstrate that the proposed method outperforms several state-of-the-art baselines. The ablation experiment proves the effectiveness of the auxiliary task of evidence veracity in fake news detection.**

Keywords*—*Multi-modal fake news; multi-task learning; external evidences; multi-level attention mechanism

I. Introduction

Social media is an important platform for people to share and obtain information, and has become an indispensable part of people's daily life. But at the same time, the characteristics of easy access and manipulation of social media information also promote the proliferation of fake news. Fake news on social media not only affects public opinion, but also does serious harm to the economy [1], politics [2], public health [3] and society. Therefore, fake news detection has become an important research issue.

The purpose of fake news detection is to automatically determine whether the statements in news posts are true or false. Some news posts contain videos or images besides words, which are more attractive and deceptive than textual news [4]. According to statistics, the average forwarding times of posts containing images are about 11 times that of posts without images [5]. Multi-modal fake news usually contains some distorted or confusing images [6]. As shown in Fig. 1, the upper image is obviously processed by tools, while the image in the lower is a misleading image that is inconsistent with the text.



Claim: A subway stop after Hurricane Sandy.
Image:

Image-related Evidence:
1. Sharks in flooded subway stop. Label: FALSE
2. There are no sharks in the subway. Label: TURE
......

Claim: Malaysia Plane(MH-370) Has Been Found.
Image:

Image-related Evidence:
1. Flight 1549 crash investigation. Label: TRUE
2. Video of the crash of Flight MH370. Label: FALSE
......

Fig. 1. Two Examples of Multi-Modal Fake News. The Claim is the Text Information of Multi-Modal News, the Image is the Visual Information Contained in the Multi-Modal News, and the Evidence is the Web Pages Extracted from Google.

Recent works have made lots of attempts on multi-modal fake news detection [7]. Some researches simply combine textual features with visual features to obtain multi-modal features [8]. Wang et al. [9] use Text-CNN and VGG-19 to extract text and image features respectively, and then simply concatenate them to classify the news. Sing et al. [10] manually design textual and visual features from four dimensions: content, organization, emotion and manipulation, and then concatenate them to detect fake news. In order to capture the interactions of multi-modal features, Wu et al. [11] stack multiple co-attention layers to fuse the multi-modal features. Qi et al. [12] extract three kinds of text-image correlations to capture multi-modal clues. However, the above methods only use the information of the news itself and neglect the use of external evidence.

To this end, this paper proposes a fake news detection method via Multi-modal and Multi-task Learning (MML). Different from previous studies, the classification of evidence veracity is used as an auxiliary task of fake news detection. MML first extracts the features of the image by a multi-layer CNN model, and then obtains evidence representations through claim-evidence correlation representation learning. Finally, the representations of image and image-related evidence are fused through the co-attention mechanism. Specifically, this paper jointly trains fake news detection and evidence classification, the two tasks share the representation of evidence.

The main contributions of the paper are as follows:

1) This paper proposes an end-to-end neural network to detect multi-modal fake news on social media by simultaneously learning the deep correlations between the image, claim and the evidence.

2) This paper extracts the image-related evidence and improves the performance of fake news detection through a multi-task learning framework.

3) Authors design detailed experiments to prove the effectiveness of the proposed model, and verify the effectiveness of multi-modal learning and multi-task learning in this task.

The remaining sections of the paper are structured as follows: Section II introduces the literature survey in the field of fake news detection and multi-task learning. After that, Section III explains the methodology. Then, Section IV describes the results and discussion followed by Section V conclusion and future enhancements.

II. Literature Survey

This section briefly summarizes the existing work in the field of fake news detection and multi-task learning.

A. Fake News Detection

For news that only contains texts, besides the text information, the propagation structure of news on social networks is commonly used to detect fake news. Liu et al. [13] presented a kernel graph attention network, which performed more fine-grained fact verification based on kernel-based attentions. Zhong et al. [14] applied semantic role labeling to parse each evidence sentence and established links between arguments to build a graph structure for information detection. Different from the graph structure constructed in the above methods, Ma et al. [15] and Bian et al. [16] modelled the propagation of posts on the Weibo platform by tree structures. Some researchers have different opinions about the research direction of fake news. They think it is very important to study the interpretability of fake news detection. Shu et al. [17] developed a joint attention graph to capture the top K interpretable sentences and user comments. Wu et al. [18] proposed a dual-view model based on collective cognition and individual cognition for interpretative claim verification.

While for multi-modal news, various methods have been proposed to utilize the multi-modal information and detect fake news. Vo et al. [19] proposed to use images as a supplement to news content, and used text matching layer and visual matching layer to detect text and images respectively. Jin et al. [20] treated each image or video as a topic and used the credibility of these topics as a new feature to detect fake news. However, a key problem with using multi-modal information for fake news detection is that the multi-modal information usually comes from another real event, and the content seems to correspond to the text in the fake news. At this point, although the image itself is real, it does not actually match the text content. The above methods ignore this problem and do not fully integrate text and multimedia content. Based on this, Wu et al. [11] proposed a joint multi-modal attention network to integrate the text features and visual features of fake news. Qi et al. [12] captured the correlation between text features and visual features by extracting three kinds of text-image features. However, these methods ignore the use of external evidence. Wen et al. [21] leveraged the semantic similarity between news and external evidence to capture the mismatch between text content and multi-modal information. However, this method did not fuse the physical features of image. To overcome the above limitations, this paper proposes a method to capture the physical features of image, and learns the deep correlations between the image, claim and the evidence.

B. Multi-Task Learning

Multi-task learning refers to the joint learning of related tasks that share representation information, so that these tasks can achieve better results than training a single task. In recent years, multi-task learning has been proved to be effective in various NLP tasks, including fake news detection. Kochkina et al. [22] constructed a multi-task learning framework consisting of three tasks: veracity classification, stance classification and rumor detection. The proposed method was represented by a shared LSTM layer (hard parameter sharing), followed by many task-specific layers. Ma et al. [23] jointly modelled rumor detection and stance classification by using two RNN-based architectures with shared layers. Wu et al. [24] explored a sharing layer of gate mechanism and attention mechanism, which can selectively capture valuable sharing features for fake news detection and stance detection. Li at al. [25] proposed a neural network model for multi-task learning of rumor detection and stance classification, including a shared layer and two task-specific layers. However, all the above multi-task learning methods are based on the joint training of fake news detection and stance detection. To the best of author's knowledge, this paper makes the first attempt to jointly model fake news detection and evidence veracity classification in multi-task learning.

III. Methodology

A. Overview

This paper proposes a Multi-modal and Multi-task Learning method for fake news detection (MML). As shown in Fig. 2, the proposed model mainly contains three parts: visual representation learning, textual representation learning, and fake news classification.

The problem definition is as follows. Suppose that $P = (p_1, \ldots, p_n)$ is a set of multi-modal news posts from social media, the text in the news is denoted as a claim C_j where $j \in [1, n]$. $E_i = \{e_i^1, e_i^2, \ldots, e_i^m\}$ is a set of evidence for news p_i, composed of the titles of web pages searched from Google. Given a news post p_i and the corresponding evidence set E_i, the main task aims to predict whether p_i is a fake news based on its multi-modal representation learned from MML. For the task of evidence veracity classification, in the training stage, this paper uses the label of evidence to learn the representation of evidence and shares it with the main task.

B. Visual Representation Learning

Since the fake-news images are often re-compressed images or tampered images, they are different from real-news

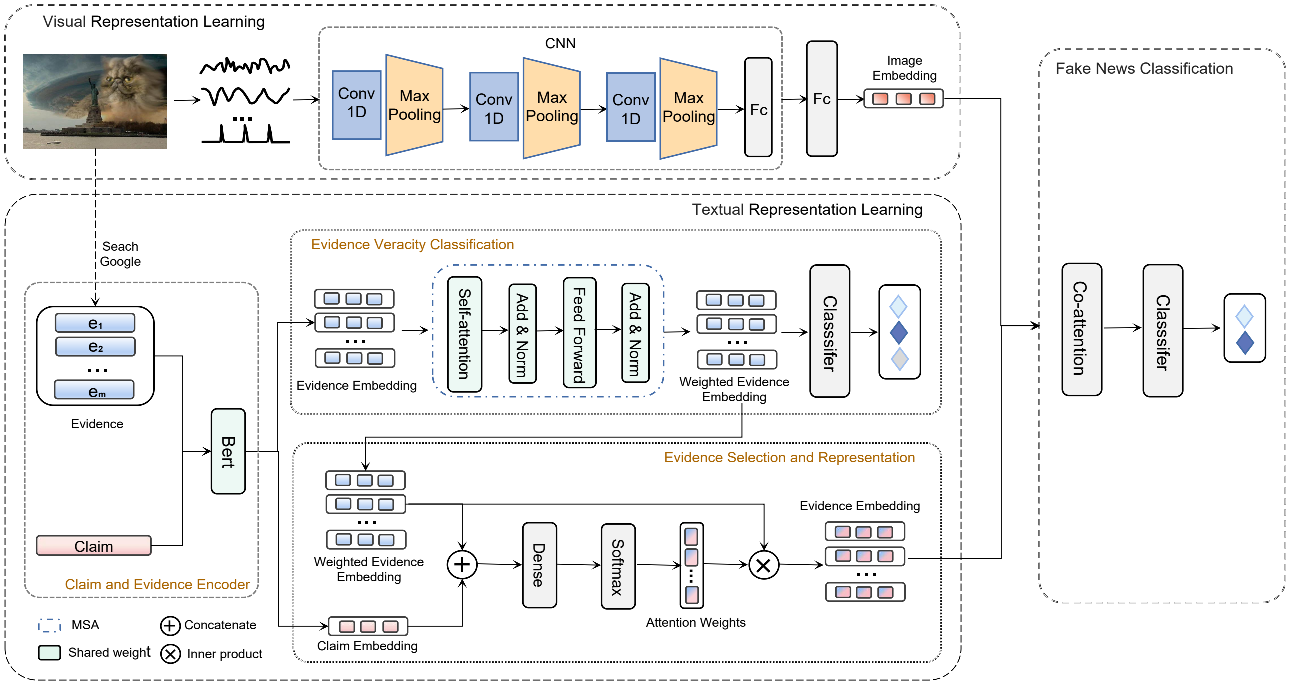


Fig. 2 Illustration of the Proposed Model MML.

images in frequency domain, which are usually periodic. Inspired by Qi et al. [26], the discrete cosine transform (DCT) is first used to transform the image in the news post from spatial domain to frequency domain, to obtain 64 hisograms, which can be represented by 64 vectors $V_0, V_1, \ldots, V_{63}$ with a fixed size. After that, this paper feeds each vector to the multi-layer CNN model consisting of three convolution blocks and a fully connected layer, where each convolution block contains a one-dimensional convolution layer and a max-pooling layer. Finally, a fully connected layer with ReLU activation function (denoted as "Fc" in Fig. 2) is added to get the feature representation of image R_v.

C. Textual Representation Learning

To capture the correlations of the semantics and visual information of the news posts, MML extracts image related web pages from Google to serve as the evidence of the claim. At the same time, in order to make a selection of evidence, MML uses the evidence veracity classification task to assist the fake news detection task. In this part, the claim and the evidence are first fed into a BERT-based encoder, then through the evidence veracity classification task, the importance of evidence is learned. Finally, the textual representation is learned based on the co-attention of claim and the evidence.

1) Claim and Evidence Encoder: This paper uses BERT to obtain the representations of claim and its corresponding m related evidences. The BERT model is a bidirectional coding representation model based on the transformer structure proposed by Devlin et al. [27]. Compared with the traditional Recurrent Neural Network (RNN) and Long Short-Term Memory networks (LSTM) used for NLP tasks, the transformer structure is more powerful in encoding texts. It consists of six encoder-decoders stacked with the same structure. Each encoder consists of two sub-layers, i.e., a feedforward layer and a multi-head attention layer, and each decoder consists of three sub-layers: a feedforward layer, a multi-head attention layer and a masked multi-head attention layer. In addition, add and normalization functions are added to each sub-layer. The BERT model achieves better performance in existing models by stacking twelve-layer Transformer Encoders.

Given a claim C and a set of evidences $E = \{e_1, e_2, \ldots, e_m\}$ corresponding to the claim, BERT model is used to generate the representations of the claim and each evidence:

$$R_c = BERT(C) \tag{1}$$

$$h_j = BERT(e_j) \tag{2}$$

where e_j is the *j*-th evidence corresponding to the claim C. Next, the total evidence representation H is obtained by concatenating the representation of each evidence:

$$H = h_1 \oplus h_2 \oplus \cdots \oplus h_m \tag{3}$$

where m represents the number of evidence corresponding to the claim C.

2) Evidence Veracity Classification: Since there are a lot of web pages searched from Google, it is of great significance for fake news detection that how to find the "useful evidences" and make use of them. Taking the evidence representation H as input, the Transformer encoder is used to capture the correlations of evidences, and a Multi-layer Perceptron (MLP) is used to classify the evidence into three pre-defined categories: True, False and Unverified.

The objective of evidence veracity classification task is to minimize the cross-entropy loss function:

$$\mathcal{L}_e = -\sum_{i=1}^{m} q_i log p_i \tag{4}$$

where p_i denotes predicted probability of evidence i, q_i refers to the ground-truth label of evidence i. By classifying the veracity of evidences, this paper can use transformer to learn the correlations of evidences and share it with the fake news detection task.

3) Evidence Selection and Representation: After obtaining the weighted evidence representations from the auxiliary task, this paper uses attention mechanism to select important evidences related to the claim. Given the claim representation R_c and evidence representation $\{h_1, h_2, \ldots, h_m\}$, this paper concatenates claim representation with each evidence representation:

$$a_j = R_c \oplus h_j \tag{5}$$

where h_j is the *j*-th evidence representation corresponding to claim.

This paper performs a linear and a softmax to calculate the attention score between the claim and the *j*-th evidence, and gains the evidence weighted representation based on claim-evidence attention R_e:

$$\alpha_j = \frac{exp(a_j W^T + b)}{\sum_j exp(a_j W^T + b)} \tag{6}$$

$$R_e = [\alpha_1 \cdot h_1, \ldots, \alpha_j \cdot h_j, \ldots, \alpha_m \cdot h_m] \tag{7}$$

where W^T denotes the weight matrix and b is the bias term, α_j is the attention score between the *j*-th evidence and the claim.

D. Fake News Classification

Given the visual representation and the textual representation, this paper uses a co-attention block to fuse the image representation R_v and image-related evidence representation R_e and obtains R'. The structure of the co-attention block is as follows:

$$R = R_e + MHA(R_e, R_v, R_v) \tag{8}$$

$$R' = R + FFN(R) \tag{9}$$

This paper feeds R' into a MLP layer to predict whether the news post is fake or not. The loss function of this part $\mathcal{L}_n$ is as follows:

$$\mathcal{L}_n = -\sum_{i}^{n} [y_i * log(\hat{y}_i) + (1 - y_i) * log(1 - \hat{y}_i)] \tag{10}$$

where y_i denotes the ground-truth label of post i and $\hat{y}_i$ indicates the predicted probability of being fake news.

The overall objective function consists of two parts: evidence classification loss and news classification loss. According to Equations 4 and 10, the objective function of MML can be defined as:

$$\mathcal{L}_{final} = \lambda \mathcal{L}_e + \mathcal{L}_n \tag{11}$$

where $\mathcal{L}_e$ denotes the evidence classification loss, $\mathcal{L}_n$ represents the news classification loss, and λ is a hyperparameter used to balance these two losses.

IV. Experimental Results and Discussion

In this section, authors conduct experiments to evaluate the effectiveness of the proposed MML. Specifically, the section aims to answer the following research questions: EQ1: Can MML improve the performance of multi-modal fake news detection? EQ2: Are visual representation and multi-task learning useful for fake news detection task? If so, how much can it improve? EQ3: Is MML model sensitive to different parameter settings?

A. Dataset

Authors evaluate the proposed MML model on the CCMR dataset, which is a multimedia fake news verification dataset with 17 events in total [21]. The dataset consists of 15,629 tweets with multimedia information, 4,625 webpages from Google and 2,506 webpages from Baidu that share similar multimedia content. Among them, this paper only uses the tweets and webpages from Google in the dataset to perform the experiment. Table I shows the statistical information of the CCMR dataset.

B. Baseline Methods

This section compares the proposed MML model with the following state-of-the-art methods:

1) **SpotFake+:** Singhal et al. [28] build a multi-modal fake news detection method based on transfer learning. The model extracts the features of text and image respectively, and then feeds the feature vectors to a fully connected layer for classification.

2) **IDM-FND:** Singhal et al. [29] develop a fake news detection framework based on inter-modality inconsistency. Firstly, the framework captures the relationship (inconsistency) among various components in news articles. Then, the features of text and image features are extracted and concatenated to detect fake news.

3) **MVNN:** Qi et al. [26] propose a multi-domain visual neural network framework, which extracts and fuses the features of frequency domain and pixel domain of images to detect fake news.

4) **MCAN:** Wu et al. [11] propose a multi-modal co-attention network to fuse the features of textual and visual features. Firstly, the network uses BERT to extract features. Secondly, the spatial domain and frequency domain features of the image are captured respectively. Finally, the multi-modal features are fused by stacking four co-attention layers.

5) **TFG:** Wen et al. [21] use cosine similarity and agreement classifiers to obtain the classification features. The network leverages the multimedia information to find the consistency and inconsistency among news from different social media platforms but sharing similar visual contents.

TABLE I. Statistics of the CCMR Dataset

ID	Event	Twitter	Google
01	Hurricane Sandy	10222	2204
02	Boston Marathon bombing	533	722
03	Sochi Olympics	274	347
04	MA flight 370	310	323
05	Bring Back Our Girls	131	108
06	Columbian Chemicals	185	63
07	Passport hoax	44	26
08	Rock Elephant	13	20
09	Underwater bedroom	113	59
10	Livr mobile app	9	15
11	Pig fish	14	20
12	Solar Eclipse	277	143
13	Girl with Samurai boots	218	60
14	Nepal Earthquake	1360	424
15	Garissa Attack	79	63
16	Syrian boy	1786	8
17	Varoufakis and zdf	61	20
	Total	15629	4625

C. Evaluating Metrics

To evaluate the performance of the proposed MML model, this paper uses four commonly used evaluation metrics: Accuracy, Precision, Recall and F1-score. Accuracy is a relatively intuitive evaluation index, which indicates the proportion of correctly classified samples in the total number of samples. Precision (P) represents the probability that the samples predicted to be true are real positive samples. Recall (R) represents the probability that positive examples in the sample are predicted to be correct. In practical evaluation of a model, both Precision and Recall should be considered, but it is difficult to compare the two values in a balanced way. The F1-score (F1) is a common method of integrating two values for evaluation:

$$F1 = \frac{2 * P * R}{P + R} \tag{12}$$

D. Implementation Details

This paper uses event 1-11 for training and event 12-17 for testing according to Wen et al. [21]. This paper sets the number of hidden layers in the Transformer encoder and the number of attention heads to 12. The maximum sequence length is set to 512. The learning rate is set to 1e-5 and the batch size is set to 8. The dropout of each layer is 0.1. The hyperparameter λ is 0.2.

E. Experimental Results and Analysis

This section compares the performance of the proposed model MML with the above baselines. From the results in Table II, authors can draw the following conclusions:

TABLE II. Results of MML Model and Baseline Models

Methods	Accuracy	Precision	Recall	F1-score
SpotFake+	0.7615	0.8212	0.7652	0.7921
IDM-FND	0.7937	0.7849	0.8231	0.8035
MVNN	0.8399	0.8173	0.8461	0.8315
MCAN	0.8573	0.8632	0.8347	0.8487
TFG	0.8912	0.8813	0.9254	0.9029
MML	**0.9225**	**0.9169**	**0.9262**	**0.9215**

TABLE III. Evaluation Results of the MML Model and Two Variants

Methods	Accuracy	Precision	Recall	F1-score
MML	**0.9225**	**0.9169**	**0.9262**	**0.9215**
MML-*w/o* Visual Representation	0.8501	0.8742	0.8439	0.8588
MML-*w/o* Evidence Veracity Classification	0.8651	0.8673	0.8426	0.8548

1) The MML model performs significantly better than all baseline models, achieving the Accuracy of 0.9225 and F1-score of 0.9215. Compared with SpotFake+ and IDM-FND, which simply combine the multi-modal features, MML achieves the greatest improvement, 16.1% in Accuracy and 12.9% in F1-score.

2) Compared with other multi-modal models MVNN and MCAN, the proposed MML model improves the Accuracy by 8.2% and 6.5%, respectively. It can be speculated that external evidence can effectively identify the correlations between text and image, and help improve fake news detection.

3) Compared with TFG, which also uses external evidence to detect fake news, MML is 3.1% and 1.8% higher in Accuracy and F1-score, respectively. This is because MML has advantages in extracting physical features of images and selecting important evidences.

F. Ablation Experiment

In this section, authors discuss the contribution of different components in the model, including visual representation learning and evidence veracity classification task. Authors remove the above two modules from MML model to obtain the following two variants: **MML- w/o Visual Representation**, which denotes MML only models the textual representation, and **MML- w/o Evidence Veracity Classification**, representing MML without multi-task learning.

The results of the two variants are shown in Table III. When the visual representation learning module is removed, the Accuracy and F1-scores drop to 0.8501 and 0.8588, respectively, showing the importance of visual representation for multi-modal fake news detection. By comparing MML with the

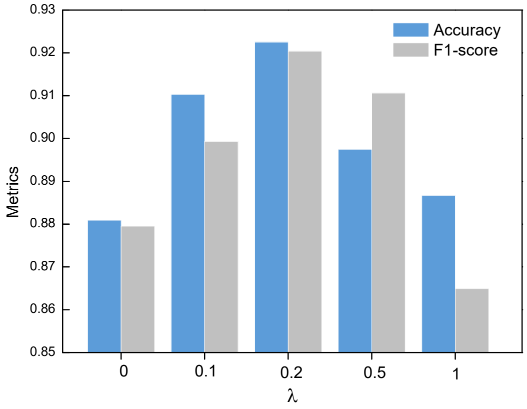


Fig. 3. Results of different Hyperparameter λ.

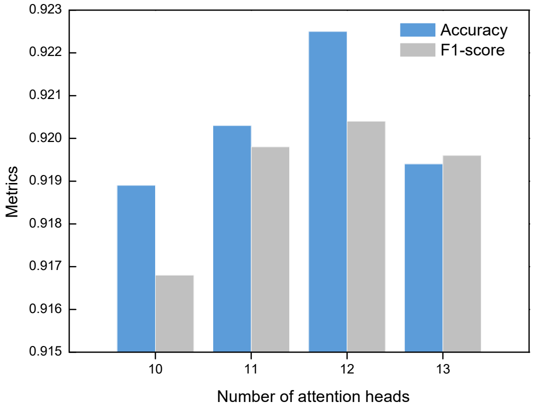


Fig. 4. Results of different Number of Attention Heads.

variant without evidence veracity classification task, authors can observe that all the evaluating metrics decreased greatly, which demonstrates the effectiveness of multi-task learning and the necessity of evidence selection and representation.

G. Parameter Sensitivity Analysis

To analyze the influence of hyperparameters on model performance, authors conduct the following two parameter sensitivity experiments.

1) Effects of hyperparameter λ: Note that λ is a weight parameter for balancing the evidence classification loss $\mathcal{L}_e$ and the news classification loss $\mathcal{L}_n$. In other words, the larger λ is, the greater the effect of evidence weight learning on fake news detection. This paper sets λ to 1, 0.5, 0.2, 0.1 and 0. The Accuracy and F1-score of different hyperparameter λ are shown in Fig. 3. Authors find that the model achieves the best performance when λ is 0.2, while the evidence classification loss brings an improvement of 5% on F1-score for the proposed MML model (λ=0, without evidence classification loss). This proves the effectiveness of the model by introducing evidence classification loss.

2) Number of attention heads: As shown in Fig. 4, authors can clearly see that the performance of the proposed model varies with the number of attention heads (i.e. 10, 11, 12 and 13). With the increase of the number of attention heads, the Accuracy and F1-score firstly increase and then decrease, and the best effect is achieved when the number of heads is 12.

V. Conclusion and Future Enhancements

This paper proposes a Multi-model Multi-task Learning model (MML) to detect multi-modal fake news on social media by modeling the image, claim and image-related evidence. By comparing MML with other competitive baseline methods, authors find that it is effective to use external evidence in this task, with an accuracy of 92.2%. In addition, besides the news classification loss, MML also introduces evidence classification loss to further optimize the model performance. By testing MML with different settings, authors observe that the proper setting of evidence classification loss can improve the performance of fake news detection. Finally, the results of the ablation experiments show that visual feature representation and evidence representation learning are beneficial to improve the fake news detection results, and the model is improved by 7.2% and 5.7%, respectively.

In the future, the authors are willing to extract the visual entity of the image and the text embedded in the image, and model them with the news text to further capture the correlation between the image and the text in multi-modal news.

Acknowledgment

Authors thank the anonymous reviewers for their constructive suggestions. This research was funded by the National Natural Science Foundation of China [61806049] and the Heilongjiang Postdoctoral Science Foundation [LBH-Z20104]. Corresponding author: Yang Li, yli@nefu.edu.cn.

References

[1] S. Kogan, T. J. Moskowitz, and M. Niessner, "Fake news: Evidence from financial markets," *Available at SSRN*, vol. 3237763, 2019, doi: 10.2139/ssrn.3237763.

[2] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017, doi: 10.1257/jep.31.2.211.

[3] A. Depoux, S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith, and H. Larson, "The pandemic of social media panic travels faster than the covid-19 outbreak," p. taaa031, 2020, doi: 10.1093/jtm/taaa031.

[4] Q. Sheng, X. Zhang, J. Cao, and L. Zhong, "Integrating pattern- and fact-based fake news detection via model preference learning," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 1640–1650, doi: 10.1145/3459637.3482440.

[5] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE transactions on multimedia*, vol. 19, no. 3, pp. 598–608, 2016, doi: 10.1109/TMM.2016.2617078.

[6] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 675–684, doi: 10.1145/1963405.1963500.

[7] Y. Fung, C. Thomas, R. G. Reddy, S. Polisetty, H. Ji, S.-F. Chang, K. McKeown, M. Bansal, and A. Sil, "Infosurgeon: Cross-media fine-grained information consistency checking for fake news detection," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 1683–1698, doi: 10.18653/v1/2021.acl-long.133.

[8] Y. Wang, F. Ma, H. Wang, K. Jha, and J. Gao, "Multimodal emergent fake news detection via meta neural process networks," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 3708–3716, doi: 10.1145/3447548.3467153.

[9] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "Eann: Event adversarial neural networks for multi-modal fake news detection," in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, 2018, pp. 849–857, doi: 10.1145/3219819.3219903.

[10] V. K. Singh, I. Ghosh, and D. Sonagara, "Detecting fake news stories via multimodal analysis," *Journal of the Association for Information Science and Technology*, vol. 72, no. 1, pp. 3–17, 2021, doi: 10.1002/asi.24359.

[11] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, "Multimodal fusion with co-attention networks for fake news detection," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2021, pp. 2560–2569. Available: https://aclanthology.org/2021.findings-acl.226.pdf.

[12] P. Qi, J. Cao, X. Li, H. Liu, Q. Sheng, X. Mi, Q. He, Y. Lv, C. Guo, and Y. Yu, "Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 1212–1220, doi: 10.1145/3474085.3481548.

[13] Z. Liu, C. Xiong, M. Sun, and Z. Liu, "Fine-grained fact verification with kernel graph attention network," *arXiv preprint arXiv:1910.09796*, 2019, doi: 10.48550/arXiv.1910.09796.

[14] W. Zhong, J. Xu, D. Tang, Z. Xu, N. Duan, M. Zhou, J. Wang, and J. Yin, "Reasoning over semantic-level graph for fact checking," *arXiv preprint arXiv:1909.03745*, 2019, doi: 10.48550/arXiv.1909.03745.

[15] J. Ma and W. Gao, "Debunking rumors on twitter with tree transformer." ACL, 2020. Available: https://ink.library.smu.edu.sg/sis_research/5599.

[16] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 549–556, doi: 10.1609/aaai.v34i01.5393.

[17] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: Explainable fake news detection," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 395–405, doi: 10.1145/3292500.3330935.

[18] L. Wu, Y. Rao, Y. Lan, L. Sun, and Z. Qi, "Unified dual-view cognitive model for interpretable claim verification," *arXiv preprint arXiv:2105.09567*, 2021, doi: 10.48550/arXiv.2105.09567.

[19] N. Vo and K. Lee, "Where are the facts? searching for fact-checked information to alleviate the spread of fake news," in *The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, 2020, doi: 10.48550/arXiv.2010.03159.

[20] Z. Jin, J. Cao, Y. Zhang, and Y. Zhang, "Mcg-ict at mediaeval 2015: Verifying multimedia use with a two-level classification model." in *MediaEval*, 2015. Available: http://ceur-ws.org/Vol-1436/Paper51.pdf.

[21] W. Wen, S. Su, and Z. Yu, "Cross-lingual cross-platform rumor verification pivoting on multimedia content," *arXiv preprint arXiv:1808.04911*, 2018, doi: 10.48550/arXiv.1808.04911.

[22] E. Kochkina, M. Liakata, and A. Zubiaga, "All-in-one: Multi-task learning for rumour verification," *arXiv preprint arXiv:1806.03713*, 2018, doi: 10.48550/arXiv.1806.03713.

[23] J. Ma, W. Gao, and K.-F. Wong, "Detect rumor and stance jointly by neural multi-task learning," in *Companion proceedings of the the web conference 2018*, 2018, pp. 585–593, doi: 10.1145/3184558.3188729.

[24] L. Wu, Y. Rao, H. Jin, A. Nazir, and L. Sun, "Different absorption from the same sharing: Sifted multi-task learning for fake news detection," *arXiv preprint arXiv:1909.01720*, 2019, doi: 10.48550/arXiv.1909.01720.

[25] Q. Li, Q. Zhang, and L. Si, "Rumor detection by exploiting user credibility information, attention and multi-task learning," in *Proceedings of the 57th annual meeting of the association for computational linguistics*, 2019, pp. 1173–1179, doi: 10.18653/v1/P19-1113.

[26] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multi-domain visual information for fake news detection," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 518–527, doi: 10.1109/ICDM.2019.00062.

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018, doi: 10.48550/arXiv.1810.04805.

[28] S. Singhal, A. Kabra, M. Sharma, R. R. Shah, T. Chakraborty, and P. Kumaraguru, "Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract)," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 10, 2020, pp. 13 915–13 916, doi: 10.1609/aaai.v34i10.7230.

[29] S. Singhal, M. Dhawan, R. R. Shah, and P. Kumaraguru, "Inter-modality discordance for multimodal fake news detection," in *ACM Multimedia Asia*, 2021, pp. 1–7, doi: 10.1145/3469877.3490614.