Fake News Detection Based on Multi-Modal Classifier Ensemble

Yi Shao
shaoyi625631@hotmail.com
School of Information Science and Engineering, Shandong Normal University
Jinan, Shandong, China

Jiande Sun*
School of Information Science and Engineering, Shandong Normal University
Jinan, Shandong, China
jiandesun@hotmail.com

Tianlin Zhang
taylorzhang19951019@gmail.com
School of Information Science and Engineering, Shandong Normal University
Jinan, Shandong, China

Ye Jiang
ye.jiang@qust.edu.cn
School of Information Science and Technology, Qingdao University of Science and Technology
Qingdao, Shandong, China

Jianhua Ma
majianhua1125@hotmail.com
School of Information Science and Engineering, Shandong Normal University
Jinan, Shandong, China

Jing Li*
lijingjdsun@hotmail.com
School of Journalism and Communication, Shandong Normal University
Jinan, Shandong, China

ABSTRACT

With the advent of the era of big data, the ubiquity of multi-modal fake news has increasingly affected information dissemination and consumption. Measurements should be taken to identify multi-modal fake news for improving the credibility of news. However, existing single-modal fake news detection models fail to detect fake news based on complete multi-modal information, while multi-modal models are often difficult to fully utilize the original information of each single modality to obtain the ultimate accuracy. To tackle above problems, we propose a novel multi-modal fake news detection method, called fake news detection based on multi-modal classifier ensemble, which takes into account the advantages of both single-modal and multi-modal models. Specifically, we design two single-modal classifiers for text and image inputs respectively. We then establish a similarity classifier to calculate the feature similarity over the modalities. We also build an integrity classifier that utilizes integral multi-modal information. Finally, all classifier outputs are integrated with an ensemble learning to increase the classification accuracy. Furthermore, we introduce the center loss, to reduce intra-class variance, which is helpful to achieve higher detection accuracy. The cross-entropy loss is used to maximize the inter-class variations while the center loss is used to minimize the intra-class variations so that the discriminative power of the learned news features can be enhanced. Experimental results on both Chinese and English datasets demonstrate that the proposed method outperforms the baseline fake news detection approaches.

KEYWORDS

Fake News Detection, Multi-Modal News, Cross-Entropy Loss, Center Loss, Cross-Media

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Yi Shao, Jiande Sun, Tianlin Zhang, Ye Jiang, Jianhua Ma, and Jing Li. 2018. Fake News Detection Based on Multi-Modal Classifier Ensemble. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

With increasingly use of social media, information dissemination has been greatly improved where multimedia sources make information sharing become much easier. However, the explosion of fake news has also caused significant harm to the global community. Fake news not only damages the credibility of information source, but also potentially cause social unrest. For example, in April 2020, fake news which related to Woody Harrelson who sharing COVID-19 conspiracy theories tied to 5G, has been widely spread on Twitter. These fake news has caused great trouble that impacted Woody himself, and has directly caused the world to panic about 5G. Consequently, the automatic fake news identification is urgently needed to minimize the impact of misinformation.

Focusing only on text and image modalities, fake news in reality is usually divided into 3 categories [11]. (1) The text is a fabrication (see Figure 1a). (2) The images themselves have been artificial tampered (see Figure 1b). (3) The image content in the news comes from an event that is related to the text but doesn't exactly match (see Figure 1c). Taking them all into account, we can draw a conclusion that both of the single-modal information and the multi-modal information should be analyzed in the multi-modal fake news detection task.

It can be seen that in the multi-modal fake news detection task, the result of fake news can sometimes be drawn directly based on only one single-modal feature, so as to avoid interference by other correct modal information. But most of existing multi-modal models ignore the importance of single-modal detection, i.e. do not fully utilize the original information of each single modality as single-modal models. To make the detection results have the advantages of single-mode detection classifiers and multi-modal ones, we propose our model. The proposed model not only utilizes the information of all modalities, but also makes an innovation on the existing multi-modal models. That is, it detects the important

(a) A picture depicting many ducks is fictitious as a "duck soldier" picture of China supporting Pakistan in eradicating the locust plague.

(b) In this picture, there is a huge fist and related giant text "China" on the snow, but one fact that it's a composite picture has been proved.

(c) This is a photo of Trump threatening to close Congress in 2019, but it is rumored to be his comeback announcement in 2022.

Figure 1: Specific cases of the three categories of multimodal fake news.

information in each single modality, which improves the utilization of original information and the accuracy of fake news detection.

The framework of the proposed model is illustrated in Figure 2. The multi-modal feature extractor obtains multi-modal feature from news posts. Then the features are fed to four independent classifiers for fake news detection. These four classifiers include two single-modal classifiers that uses text feature and image feature to detect fake news respectively which can determine whether the text or image itself has been fabricated or tampered with. Our method adopts VGG-19 [12] and TEXT-CNN [13] to obtain body text and images features respectively. The other two classifiers are the similarity classifier and the integrity classifier. The similarity classifier aims to distinguish the authenticity of news based on the similarity between images and text. And the addition of the integrity classifier is inspired by existing multi-modal detection methods. The integrity classifier is designed to identify fake news through the text feature as well as the image feature. The above classifiers are combined with weights that can be adjusted automatically during the training procedure. Another innovation of the model is the combination of the traditional cross-entropy loss and center loss [14], which is proved to be helpful for the optimization of the model. Finally, the classification results of the four classifiers are weighted to obtain the final classification results.

The main contributions of our method are summarized as follows:

- The proposed model contains two single-modal classifiers and two multi-modal classifiers respectively, which can not only directly detect single-modal information that may be tampered with, but also take into account the advantages of comprehensive consideration of multi-modal classifiers. And they are trained to get an ensemble fake news detector which gets higher accuracy by automatically updated weights.
- The proposed model adopts the joint optimization of center loss and cross-entropy loss. The idea of adding center loss was inspired by face recognition algorithms. Reducing the center loss can reduce the intra-class variance and make the features learned by the model more discriminative. With the joint optimization of cross-entropy loss and center loss, the proposed model can obtain higher accuracy.
- We have made an approximately 20% augmentation to the Weibo fake news dataset created by previous researchers [15] (the dataset will be described in detail in the Section 4), and conducted experiments on both Chinese Weibo dataset and English Fakeddit dataset. Experiment results on these datasets show that the proposed method outperforms other state-of-the-art methods.

The rest of the paper is organized as follows. Related works is summarized in Section 2. The proposed method is described in Section 3. Experimental results are shown in Section 4. Conclusion is made in Section 5.

2 RELATED WORKS

In this section, we briefly review the work related to the proposed model. We mainly focus on the following two topics: fake news detection and center loss function.

2.1 Fake News Detection

Fake news detection has become an active research topic. Various techniques have been proposed for it, which can be roughly categorized into two categories, i.e., single-modal based and multi-modal based methods.

2.1.1 Single-modal based Fake News Detection. In single-modal based methods, single type of, often textual or visual, information such as contents, profiles and descriptions, or semantic contents and resolutions of the pictures are to detect fake news automatically.

Recently, the importance of text features has been proven by many literatures [18, 19, 20, 21] of fake news detection. Fake news detection focusing on text, which can be roughly divided into linguistic features based methods [1, 2], deception modeling base methods [3, 4], clustering based methods [3], predictive modeling based methods [3] and content cues based methods [5]; more and more researchers concentrate on analyzing various modal information [6], [7]. In [8], Steinebach et al. aims to recognize photo-montages

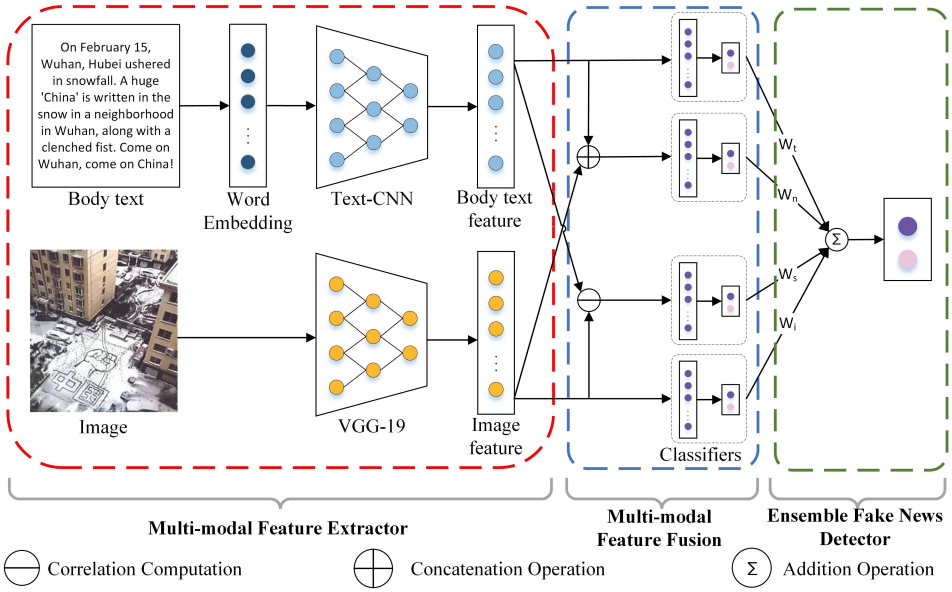

Figure 2: The model framework. The network in red dotted frame is the Multi-modal Feature Extractor, the network in blue dotted frame is the Multi-modal Feature Fusion, and the network in green dotted frame is the Fake News Detector.

automatically based on feature detection. In [9], Chen et al. proposed detect fake views about online video-on-demand services. In [10], Todisco et al. focuses on the automatic detection of spoofing attacks which can threaten the reliability of automatic speaker verification. There is also a method proposed by Ma et al. [21], which deploys a recurrent neural network to learn posts in time series as a representation of text features, proving the model based on deep learning is more effective than traditional hand-crafted algorithms.

Simultaneously, visual features have been proven to be another important indicator in fake news detection. The researches show that news readers pay more attention to visual content than to textual content, which leads to a strong guiding effect of visual content on people [22, 23]. This also proves that the detection of visual features is irreplaceable in the tasks of fake news detection. Undoubtedly, there are many studies on single-modal fake news detection models that focus on visual features [24, 25, 26, 27].

However, this category of methods can only detect fake news in which the text or image content itself has been fabricated or tampered with due to the natural defect of these methods that them can only detect fake news by the feature of a specific modality.

2.1.2 Multi-modal based Fake News Detection. In the era of big data, the content form of news tends to coexist with multiple modal information, and provides potential to many multi-modal based fake news detection methods with deep neural network. With the inputs of multi-modal features, specially designed deep neural networks obtain the ability to achieve excellent performance in cross-modal tasks such as visual question answering [16], and, of course, fake news classification [15].

However, the multi-modal fake news detector relies too much on the comprehensive expression of news content in multiple modalities, i.e., this category of method often fails to obtain the detection results that could be obtained directly from a single modal feature.

In order to overcome the respective limitations of the above two types of detection methods and take the advantages of both, we propose a novel model which is an ensemble of different single-modal based fake news detectors and multi-modal based ones. In the proposed model, a weighted combination of single-modal classifiers and multi-modal classifiers is realized, and the weights can be automatically updated during the training process.

2.2 Center Loss Function

Inspired by the paper [3] in the field of face recognition, we incorporate the special error function proposed in it, the center loss function, into our model. The center loss function reduces the intra-class variability of each class by finding the classification center of each class, so that the final classification accuracy becomes higher.

We adopt the center loss method in the proposed model to make the two classifications more discriminative, which leads to further improvement in classification accuracy. Specifically, the traditional cross-entropy loss function and the center loss function are linearly combined, and their coefficients are α and 1-α, respectively. We use as a hyper-parameter and show the effect of different values on model performance in section 4.3.

3 PROPOSED METHODS

In this section, we first introduce the three components of the proposed model, which are the multi-modal feature extractor, the multi-modal feature fusion, and the fake news detector, and then describe how to integrate these three components to learn the

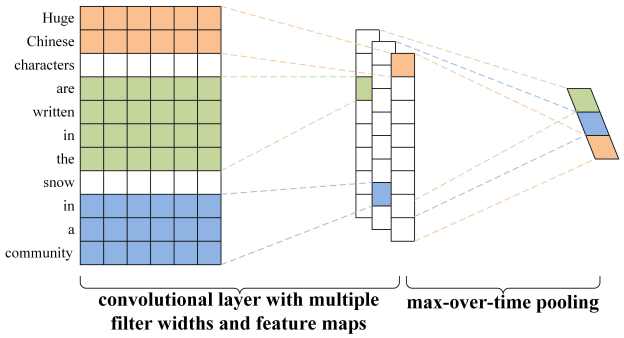

Figure 3: The structure of Text-CNN.

transferable feature representations. The proposed multi-modal fake news detection method can be divided into three parts as is shown in Figure 2.

3.1 Multi-modal Feature Extractor

Multi-modal feature extractor is responsible for the task of extracting the informative features from both textual and visual content. It is the fundamental step of the detection work. The multi-modal feature extractor is divided into two parts, which are for textual and visual content respectively.

Suppose that $\Omega = \{\omega_j\}_{j=1}^{m}$ is a social media dataset with m instances, $\omega_j = (v_j, t_j, l_j)$, where v_j is the raw pixels of image j, t_j is the body text related to the j-th image, and $l_j \in \{0, 1\}$ is authenticity label assigned to ω_j. If belongs to real news class, l_j = 0. Otherwise, l_j = 1.

3.1.1 Textual Feature Extractor. The input of the text feature extractor is the raw body text t_j related to image v_j. The core of textual extractor is a Text-CNN [2] model. Text-CNN applies convolutional neural network to the text classification task by using multiple filters with different sizes to extract the key information, which can better capture the local correlation in the sentence.

Specifically, converting each word in text t_j into a word embedding should be achieved at first. To the i-th word in text t_j, its corresponding embedding is $E_i \in R^k$, where k is the dimension of each embedding. Following that, we can obtain the complete representation of the text t_j:

$$E_{1:n} = E_1 \oplus E_2 \oplus ... \oplus E_n \tag{1}$$

where $\oplus$ is the concatenation operator. For a convolution kernel of size h, assuming that the contiguous sequence of h words generated from the i-th word is S_i, the convolution operation used to obtain S_i can be represented as:

$$S_i = \sigma(W_{filter} \cdot E_{i:i+h-1}) \tag{2}$$

where $\sigma(\bullet)$ is the ReLU activation function and W_{filter} is the weight of this kernel. Every S_i is a scalar value, and i = 1, 2, ..., $n - h + 1$.

After the max-pooling operation on these scalar values, the final operation result of this kernel for text t_j can be represented as S. Assuming there are n_h different kernel sizes, and there are n_{filter} kernels of each size, the feature vector of text t_j obtained by the entire text extractor is $R_j^T \in R^{n_h \cdot n_{filter}}$. We also add a fully connected layer behind Text-CNN to keep the dimension of the output same as visual features:

$$F_j^T = \sigma(W_T \cdot R_j^T) \tag{3}$$

where $F_j^T \in R^c$. Hence, the output of the text feature extractor for the text t_j can be expressed as:

$$F_j^T = f_T(t_j; \theta_t) \tag{4}$$

where θ_t is the denotes all the parameters of textual feature extractor, and the textual features obtained from all posts can be uniformly expressed as $F^T \in R^{c \cdot m} = \{F_j^T\}_{j=1}^{m}$.

3.1.2 Visual Feature Extractor. We obtain the feature of v_j by an eight-layer convolutional neural network, which consists of the pre-trained VGG-19 and is followed by a fully connected layer. Let θ_v denotes the parameters of VGG-19, the learned image feature F_j^V can be represented as follows:

$$F_j^V = f_V(t_j; \theta_v) \tag{5}$$

where $F_j^V \in R^c$, and the visual features obtained from all posts can be uniformly expressed as $F^V \in R^{c \cdot m} = \{F_j^V\}_{j=1}^{m}$.

3.2 Multi-modal Feature Fusion

In this part, the multi-modal features are fused and fed into two single-modal classifiers and two multi-modal classifiers respectively.

The classifiers include the text classifier, the image classifier, the similarity classifier and the integrity classifier. The text classifier is built by two fully connected layers, denoted as C_t. And the probability of F_j^T being a fake one can be obtained as follows:

$$P_{C_t} = C_t(F_j^T; \theta_{C_t}) \tag{6}$$

where θ_{C_t} denotes the parameters of C_t.

Similarly, a two-layer fully-connected layer of the image classifier is denoted as C_i, the probability of F_j^V being a fake one can be obtained as follows:

$$P_{C_i} = C_i(F_j^V; \theta_{C_i}) \tag{7}$$

where θ_{C_i} denotes the parameters of C_i.

As we all know, if the body text/image is fake in a news, the news is fake obviously. Thus F_j^T or F_j^V being a fake one is equal to ω_j being fake.

Our work concentrates on fake news detection according to the content of the news. Hence, in order to fully utilize the given multi-modal information, we build the similarity classifier and the integrity classifier. The similarity classifier is represented as C_s, which includes two fully-connected layers, and so is the integrity classifier. C_s is constructed based on the truth that some fake news is inconsistent with images and body texts. Therefore, we calculate the cosine similarity between F_j^T and F_j^V, which is fed into C_s to obtain the probability of ω_j being a fake one, denoted as P_{C_s}:

$$P_{C_s} = C_s(cos(F_j^V, F_j^V); \theta_{C_s}) \tag{8}$$

where θ_{C_s} denotes the parameters of C_s, cos(x, y) denotes the cosine similarity between x and y.

Table 1: Statistics of datasets.

Statistics		Weibo	Fakeddit
Training Set	Fake News	5568	5854
	Authentic News	3724	4982
Testing Set	Fake News	1591	1672
	Authentic News	1064	1432
Validation Set	Fake News	795	836
	Authentic News	532	712
All		13274	15479

We use C_n to denote the image and text integrity classifier, which aims to identify fake news based on the integral multi-modal information. F_j^T and F_j^V are concatenated as the input of C_n, and the probability of ω_j being a fake one can be calculated as follows:

$$P_{C_n} = C_n(F_j^V \oplus F_j^V; \theta_{C_n}) \tag{9}$$

where θ_{C_n} denotes the parameters of C_n. Benefiting from modifying the text feature to the same dimension as the visual feature in feature extractor, here we can directly concatenate F_j^T and F_j^V.

To integrate the four different classifiers into a unified framework, we set four weights denoted as W_t, W_i, W_s, W_n, and the final probability of ω_j being fake can be given as follows:

$$P = W_t \cdot P_{C_t} + W_i \cdot P_{C_i} + W_s \cdot P_{C_s} + W_n \cdot P_{C_n} \tag{10}$$

The weights can be adjusted automatically during the training procedure according to the loss of corresponding classifier. Specifically, we initialize $W_t = W_i = W_s = W_n = 0.25$ and then normalize the loss of the four classifiers, finally we average the weights and the reciprocal of loss to get the updated weights.

3.3 Fake News Detector

After the final probability P obtained in Eq.(10), a softmax function is added to achieve the final fake news classification.

In order to measure the distance between the predicted label and the ground-truth more clearly, we adopt cross-entropy loss to calculate the detection loss as follows:

$$\min_{\theta_{C_t},\theta_{C_i},\theta_{C_s},\theta_{C_n}} L_1 = -E_{\omega\sim\Omega}[l \log Pr + (1-l)\log(1-P)] \tag{11}$$

To further enhance intra-class compactness, we also introduce center loss, which can be calculated as follows:

$$\min_{\theta_{C_t},\theta_{C_i},\theta_{C_s},\theta_{C_n}} L_2 = \frac{1}{2}\Sigma_{j=1}^{m} \parallel P_j - l_j \parallel_2^2 \tag{12}$$

Center loss can be regarded as an auxiliary loss function, which is used to minimize the intra-class differences. The discrimination of the learned features will be higher in the way of combining cross-entropy loss and center loss. The overall objective function can be deduced as follows:

$$\min_{\theta_{C_t},\theta_{C_i},\theta_{C_s},\theta_{C_n}} L = \alpha L_1 + (1-\alpha)L_2 \tag{13}$$

where α is a hyper-parameter used to balance the two, and we will compare the effect of its value on the experimental results in section 4.2.

4 EXPERIMENTS

4.1 Datasets

In order to adapt to the differences in various cultures, we conducted experiments on Chinese and English datasets, respectively.

4.1.1 Weibo Dataset. Weibo Dataset is created by [15] for Chinese fake news detection. The authentic news is collected from Xinhua News Agency, the biggest and most influential media organization in China. The fake news is collected from Weibo, a Chinese microblogging website. They are crawled from the official rumor debunking system of Weibo covering from May, 2012 to January, 2016. There are a total of 13,274 items in the Weibo Dataset, including 7,954 fake news and 5,320 authentic news.

4.1.2 Fakeddit Dataset. Fakeddit Dataset [30] is a novel multimodal dataset consisting of over 1 million samples from multiple categories of fake news and corresponding authenticity labels. The dataset is collected from 2019 and is still being updated today. There are a total of 15,479 items in the Fakeddit dataset, including 8,362 fake news and 7,112 real news. The selected data in both datasets are news with fluent text and pictures containing a lot of useful information. Besides, we preprocess the dataset similar to attRNN [15], and the whole datasets are split into the training, validation, and testing sets in a 7:1:2 ratio.

The detailed statistics of two datasets are listed in Table 1.

4.2 Performance Comparison

We compared the proposed model with these traditional single-modal ones, including single textual modal detection methods, single visual modal detection method, and multi-modal ones.

4.2.1 Single-modal detection model. We embed each post text into a 400-dimension paragraph embedding feature and then feed them to a logistic classifier to train a single textual model. Meanwhile, the 4096-dimension features got from a pre-trained VGG net are used to train a logistic regression model which is regarded as the single visual model.

4.2.2 Multi-modal detection model. **VQA** [5] means Visual Question Answering, which aims to provide an accurate natural language answer about the given image. For a fair comparison, we change VQA model, a multiclass classification, to a binary classification one by replacing the final multi-class layer with a binary-class layer, and the LSTM layer of VQA is changed to one layer. The modified algorithm is denoted as VQA*. **NeuralTalk** [6] is proposed for image caption by using deep recurrent framework. We obtain news representations by averaging the output of the RNN at each time step, which are then fed into a fully connected layer followed by an entropy loss layer to make predictions. **att-RNN** [15] adopts attention module to identify fake news based on the textual, visual and social context features. In our experiments for a fair comparison, we remove the social context processing part, but the remaining parts are set to be same. **EANN** [28] consists of three main components: the multimodal feature extractor, the fake news detector and the event discriminator. The multimodal feature extractor extracts textual and visual features from posts. It works with the fake news detector to learn the discriminative representation for detection of fake news. The event discriminator is responsible for removing

Table 2: Performance of the proposed method and baseline single-modal/multi-modal feature based methods.

Dataset	Method	Accuracy	Fake News			Authentic News		
			Precision	Recall	F1	Precision	Recall	F1
Weibo	Textual	0.592	0.605	0.531	0.566	0.581	0.653	0.615
	Visual	0.608	0.610	0.605	0.607	0.607	0.611	0.609
	VQA*[5]	0.736	0.797	0.634	0.706	0.695	0.838	0.760
	Neural Talk[6]	0.726	0.794	0.713	0.692	0.684	0.840	0.754
	att-RNN[4]	0.772	0.854	0.656	0.742	0.720	**0.889**	0.795
	EANN[28]	0.782	0.827	0.697	0.756	0.752	0.863	0.804
	MVAE[29]	0.824	**0.854**	0.769	0.809	0.802	0.875	**0.837**
	ours	**0.824**	0.835	**0.845**	**0.840**	**0.812**	0.832	0.822
Fakeddit	Textual	0.507	0.555	0.556	0.555	0.467	0.528	0.496
	Visual	0.596	0.695	0.518	0.594	0.524	0.633	0.573
	VQA*[5]	0.631	0.712	0.512	0.596	0.590	0.693	0.637
	NeuralTalk[6]	0.612	0.698	0.610	0.651	0.612	0.712	0.658
	att-RNN[4]	0.745	0.798	0.637	0.708	0.627	0.713	0.667
	EANN[28]	0.699	0.750	0.628	0.684	0.648	0.720	0.682
	MVAE[29]	0.784	0.789	0.699	0.741	0.702	0.717	0.709
	ours	**0.804**	**0.838**	**0.749**	**0.791**	**0.704**	**0.728**	**0.716**

any event-specific features. It is also possible to detect fake news using only two components, the multimodal feature extractor and the fake news detector. Hence, for a fair comparison, in our experiments, we work with a variant of EANN which does not include the event discriminator. **MVAE** [29] contains a variational autoencoder which is capable of learning probabilistic latent variable models by optimizing a bound on the marginal likelihood of the observed data. We compare the above method with the proposed model on Weibo Dataset and Fakeddit Dataset respectively.

As shown in Table 2, it can be found that our method can identify fake news more accurately. Since the single-modal baseline model cannot make a comprehensive judgment based on the information of all modalities, the detection accuracy on the two datasets is significantly lower than that of the baseline multi-modal model and the proposed model.

The multimodal baseline model, starting with VQA*, performs comprehensively on both datasets and significantly outperforms both unimodal baselines. But from Table 2, EANN performs poorly on the Fakeddit dataset. This is because the news data in the Fakeddit dataset often does not have strict domain classification, which makes EANN, which relies on a unique adversarial network mechanism to eliminate event-specific features, unable to exert its maximum effect.

The proposed model not only outperforms the single-modal baseline model, but also obtained better performance than the multi-modal baseline detection method, which is because our method can effectively maintain the authenticity and discrimination of news during feature mapping by retaining the original single-modal information and directly judge the authenticity of the news based on the original important information.

Table 3: The highest average accuracy on two datasets of each ablation model and the full model.

model	Weibo	Fakeddit
C_t	0.755	0.689
C_i	0.704	0.711
C_s	0.764	0.694
C_n	0.770	0.696
$C_t + C_i$	0.795	0.754
$C_t + C_s$	0.764	0.732
$C_t + C_n$	0.759	0.780
$C_i + C_s$	0.776	0.752
$C_i + C_n$	0.780	0.730
$C_s + C_n$	0.758	0.764
$C_t + C_i + C_s$	0.794	0.800
$C_t + C_i + C_n$	0.810	0.752
$C_t + C_s + C_n$	0.810	0.794
$C_i + C_s + C_n$	0.803	0.785
$C_t + C_i + C_s + C_n$	**0.821**	**0.807**

4.3 Ablation Analysis and Hyper-parameter Analysis

The performance of different ablation models varies with different values of the hyper-parameter α. That is to say, it is necessary to analysis and compare the performance of each ablation model under different α values.

4.3.1 Ablation Analysis. The combinations of proposed four classifiers can be further divided into one-classifier models, two-classifier models and three-classifier models. The one-classifier models are

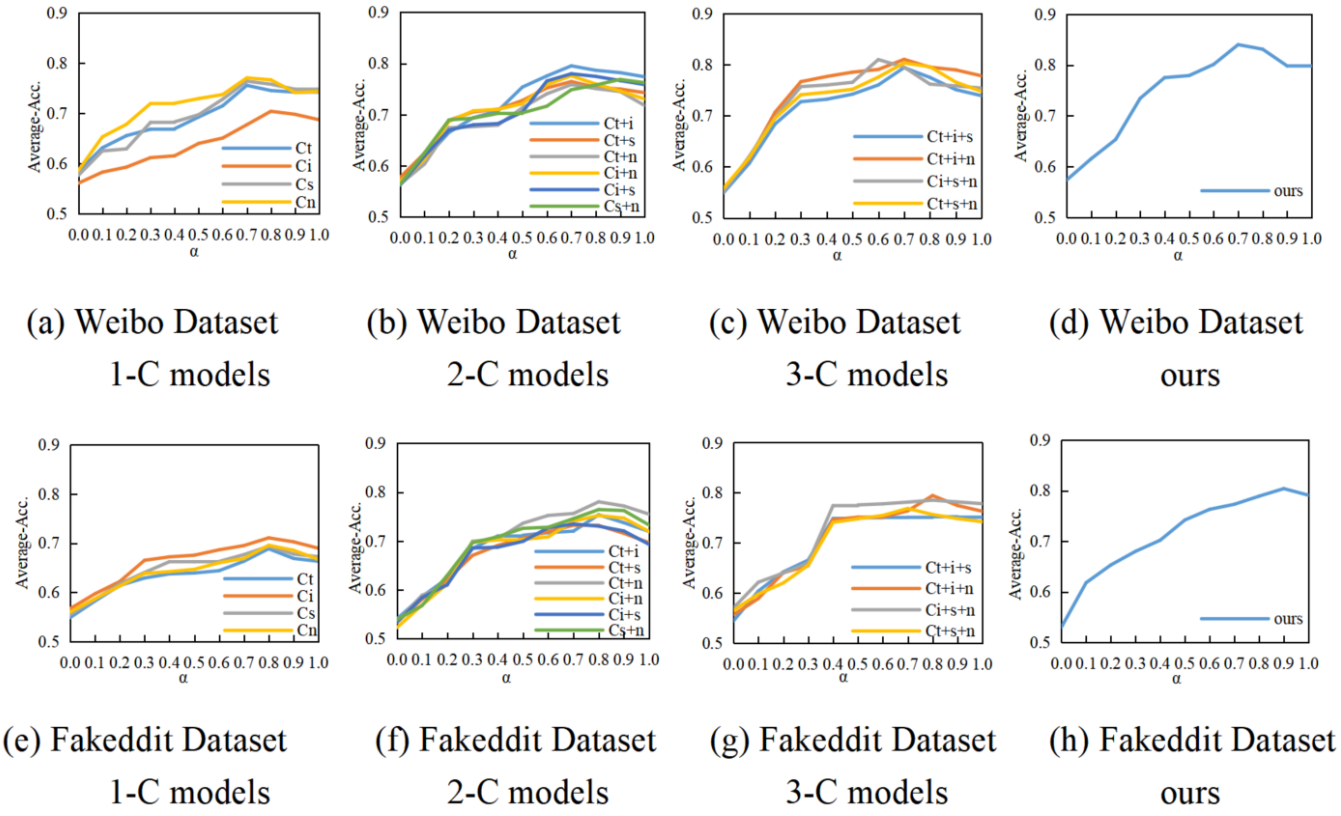

Figure 4: Average accuracy of ablation models with different values of hyper-parameter in detecting fake and authentic news on two datasets.

obtained by choosing one classifier from the four, including C_t model, C_i model, C_s model and C_n model. The two-classifier models include $C_t + C_i$ model, $C_t + C_s$ model, $C_t + C_n$ model, $C_i + C_s$ model, $C_i + C_n$ model, and $C_s + C_n$ model. We abandon one classifier from the four to construct the three classifier models, including $C_t + C_i + C_s$ model, $C_t + C_i + C_n$ model, $C_t + C_s + C_n$ model and $C_i + C_s + C_n$ model. Here we call the above models as ablation models.

We set α to take values between 0.0 and 1.0 every 0.1, and calculate the average detection accuracy of all news in each dataset for each ablation model under these α values. The results are shown in Figure 4.

It is obvious from Figure 4(a) that the pure visual feature classifier C_i has the worst performance on the Weibo dataset, because the pictures in Weibo fake news are often not fabricated but replaced from other real news, which makes the pure visual feature classifier less effective. And the number of pictures in each Weibo news post is usually small. The above two reasons make it difficult for a single visual modality method to detect multimodal fake news from Weibo. However, judging from the performance of the following relatively complete classifier combinations, our ensemble method can make up for the shortcomings of the single classifiers to a certain extent. The performance of the pure visual feature classifier on the Fakeddit dataset is much better, because a large proportion of the fake news in this dataset uses forged pictures, which makes the visual feature classifier better than the other three. While the full proposed model performs best on both datasets. Based on the above, it is proved that the proposed model is correct for direct classification using a unimodal classifier in special cases.

We can also draw a conclusion from Figure 4 that when the number of classifiers increases, the classification performance of the model is roughly improved. Two-classifier models achieve better performance than single classifier models, while three-classifier models achieve better performance than two-classifier models, and our model with four classifiers performs better than its variants with less classifiers, and finally, the points with the highest average accuracy on both datasets appear on the curves of the full model, which indicates that the full model outperforms the ablation model. For clarity, we plotted histograms of the highest average accuracy for each ablation model and the full model, as shown in Table 3. This confirms that each classifier is an essential component in the proposed model.

Thus, we can draw a conclusion that we obtain a stronger classifier based on ensemble learning by combining different classifiers into a unified framework.

4.3.2 Hyper-parameter Analysis. A conclusion can be drawn from Figure 4 that when α is small, the accuracy of all models is terribly low. This is because when the value of α is small, the error

function is equivalent to a pure center loss. At this time, the error function loses the effectiveness of the cross entropy loss to reduce the classification error, and the center loss will continue to compact and solidify the existing classification, and the final accuracy will be maintained at about 50% after initialization. That is to say, the proportion of the center loss cannot be too large, otherwise the detection effect of the model will be much worse. In classification problems, center loss is not a substitute for cross-entropy loss.

When the value of α gradually increases from 0, the accuracy gradually increases. This is because the proportion of the cross-entropy loss gradually increases, and the error function gradually shows the effectiveness of the cross-entropy loss in reducing the classification error. Fake news detection is essentially a classification problem, and cross-entropy loss is a very classic choice in the optimization process of classification models.

The accuracy of the model peaks when α reaches a threshold. For the full model, this threshold is around 0.7 (on the Weibo Dataset) and 0.9 (on the Fakeddit Dataset). At this time, the adjustment effect of center loss on cross entropy loss is just right. When the center loss has a small proportion, it can not only play the role of the center loss to make the classification more compact, but also will not affect the loss too much and make the loss function lose the effect of the cross entropy loss. Our goal is to have the error function do the best of both worlds like this, so that the accuracy peaks.

When the value of α reaches a large value, the accuracy rate will decrease slightly if α continues to increase. This is because when α is too large, the error function will degenerate into an ordinary cross-entropy loss, and the adjustment effect of the center loss will not be available at this time. When the proportion of center loss is 0, the accuracy rate is lower than the peak value, indicating that the center loss does have an additional benefit in the classification problem. This proves the feasibility of adding center loss to the error function in order to improve the accuracy.

When the value of α is very large, the peak of the accuracy can be reached, and when the threshold is exceeded, the decrease of the accuracy is not obvious. The insignificant downward trend shows that the center loss actually only plays a role in adjusting and correcting the final error function, rather than a decisive role. As mentioned above, it is the cross entropy loss that plays a decisive role in the optimization process of the classification problem. But the process of decreasing the accuracy as the value of α increases, indeed confirms the positive role of center loss in fake news detection.

5 CONCLUSION

In this paper, we propose a novel fake news detection methods based on multi-modal classifier ensemble. In order to fully utilize the given information, we train four different classifiers into an ensemble fake news detector, including the text classifier, the image classifier, the similarity classifier and the integrity classifier. The weights of these weak classifiers are updated automatically, which is related to the prediction accuracy of the previous epoch. Besides, we combine cross-entropy and center loss into the training procedure to improve the discrimination of classifiers. We also discuss the variation in the average accuracy of individual ablation models and the full model with different proportions of center loss, thereby exploring the effect and best ratio of center loss while proving that each classifier of the proposed model is indispensable. Extensive experiments on the experimental datasets demonstrate that our idea of integrating single-modal classifiers with cross-modal ones can indeed improve accuracy, and can outperform existing single-modal baseline fake news detection models and some mainstream multi-modal baseline models.

6 ACKNOWLEDGMENTS

This work was supported in part by Scientific Research Leader Studio of Jinan (No. 2021GXRC081), and in part by Joint Project for Smart Computing of Shandong Natural Science Foundation (ZR2020LZH015).

REFERENCES

[1] Pérez-Rosas V, Kleinberg B, Lefevre A, et al. Automatic detection of fake news[J]. arXiv preprint arXiv:1708.07104, 2017.
[2] Conroy N K, Rubin V L, Chen Y. Automatic deception detection: Methods for finding fake news[J]. Proceedings of the association for information science and technology, 2015, 52(1): 1-4.
[3] Rubin V L, Conroy N J, Chen Y. Towards news verification: Deception detection methods for news discourse[C]//Hawaii International Conference on System Sciences. 2015: 5-8.
[4] Rubin V L, Lukoianova T. Truth and deception at the rhetorical structure level[J]. Journal of the Association for Information Science and Technology, 2015, 66(5): 905-917.
[5] Chen Y, Conroy N J, Rubin V L. Misleading online content: recognizing clickbait as" false news"[C]//Proceedings of the 2015 ACM on workshop on multimodal deception detection. 2015: 15-19.
[6] Yu E, Sun J, Li J, et al. Adaptive semi-supervised feature selection for cross-modal retrieval[J]. IEEE Transactions on Multimedia, 2018, 21(5): 1276-1288.
[7] Wang L, Zhu L, Yu E, et al. Fusion-supervised deep cross-modal hashing[C]//2019 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2019: 37-42.
[8] Steinebach M, Gotkowski K, Liu H. Fake News Detection by Image Montage Recognition[C]//Proceedings of the 14th International Conference on Availability, Reliability and Security. 2019: 1-9.
[9] Chen L, Zhou Y, Chiu D M. Fake view analytics in online video services[C]//Proceedings of Network and Operating System Support on Digital Audio and Video Workshop. 2014: 1-6.
[10] Todisco M, Delgado H, Evans N W D. A new feature for automatic speaker verification anti-spoofing: constant q cepstral coefficients[C]//Odyssey. 2016, 2016: 283-290.
[11] Cao J, Qi P, Sheng Q, et al. Exploring the role of visual content in fake news detection[J]. Disinformation, Misinformation, and Fake News in Social Media, 2020: 141-161.
[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
[13] Chen Y. Convolutional neural network for sentence classification[D]. University of Waterloo, 2015.

[14] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition[C]//European conference on computer vision. Springer, Cham, 2016: 499-515.
[15] Jin Z, Cao J, Guo H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]//Proceedings of the 25th ACM international conference on Multimedia. 2017: 795-816.
[16] Antol S, Agrawal A, Lu J, et al. Vqa: Visual question answering[C]//Proceedings of the IEEE international conference on computer vision. 2015: 2425-2433.
[17] Xue J, Wang Y, Tian Y, et al. Detecting fake news by exploring the consistency of multimodal data[J]. Information Processing & Management, 2021, 58(5): 102610.
[18] Ahmed B, Ali G, Hussain A, et al. Analysis of Text Feature Extractors using Deep Learning on Fake News[J]. Engineering, Technology & Applied Science Research, 2021, 11(2): 7001-7005.
[19] Mangal D, Sharma D K. Fake news detection with integration of embedded text cues and image features[C]//2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2020: 68-72.
[20] Faustini P H A, Covoes T F. Fake news detection in multiple platforms and languages[J]. Expert Systems with Applications, 2020, 158: 113503.
[21] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks[J]. 2016.
[22] Vinyals O, Toshev A, Bengio S, et al. Show and tell: A neural image caption generator[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3156-3164.
[23] You Q, Cao L, Jin H, et al. Robust visual-textual sentiment analysis: When attention meets tree-structured recursive neural networks[C]//Proceedings of the 24th ACM international conference on Multimedia. 2016: 1008-1017.
[24] Mangal D, Sharma D K. Fake news detection with integration of embedded text cues and image features[C]//2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2020: 68-72.
[25] Wang Y, Qian S, Hu J, et al. Fake news detection via knowledge-driven multimodal graph convolutional networks[C]//Proceedings of the 2020 International Conference on Multimedia Retrieval. 2020: 540-547.
[26] Kumari R, Ekbal A. Amfb: Attention based multimodal factorized bilinear pooling for multimodal fake news detection[J]. Expert Systems with Applications, 2021, 184: 115412.
[27] Xue J, Wang Y, Tian Y, et al. Detecting fake news by exploring the consistency of multimodal data[J]. Information Processing & Management, 2021, 58(5): 102610.
[28] Wang Y, Ma F, Jin Z, et al. Eann: Event adversarial neural networks for multi-modal fake news detection[C]//Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining. 2018: 849-857.
[29] Khattar D, Goud J S, Gupta M, et al. Mvae: Multimodal variational autoencoder for fake news detection[C]//The world wide web conference. 2019: 2915-2921.
[30] Nakamura K, Levy S, Wang W Y. r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection[J]. arXiv preprint arXiv:1911.03854, 2019.