Article

Collaborative Mixture-of-Experts Model for Multi-Domain Fake News Detection

Jian Zhao [1,2,†], Zisong Zhao [1,†], Lijuan Shi [3], Zhejun Kuang [2,*] and Yazhou Liu [1]

[1]   College of Cyber Security, Changchun University, Changchun 130022, China
[2]   College of Computer Science and Technology, Changchun University, Changchun 130022, China
[3]   College of Electronic Information Engineering, Changchun University, Changchun 130022, China
*   Correspondence: kuangzhejun@ccu.edu.cn
†   These authors contributed equally to this work.

Abstract: With the widespread popularity of online social media, people have come to increasingly rely on it as an information and news source. However, the growing spread of fake news on the Internet has become a serious threat to cyberspace and society at large. Although a series of previous works have proposed various methods for the detection of fake news, most of these methods focus on single-domain fake-news detection, resulting in poor detection performance when considering real-world fake news with diverse news topics. Furthermore, any news content may belong to multiple domains. Therefore, detecting multi-domain fake news remains a challenging problem. In this study, we propose a multi-domain fake-news detection framework based on a mixture-of-experts model. The input text is fed to BertTokenizer and embeddings are obtained by jointly calling CLIP to obtain the fusion features. This avoids the introduction of noise and redundant features during feature fusion. We also propose a collaboration module, in which a sentiment module is used to analyze the inherent sentimental information of the text, and sentence-level and domain embeddings are used to form the collaboration module. This module can adaptively determine the weights of the expert models. Finally, the mixture-of-experts model, composed of TextCNN, is used to learn the features and construct a high-performance fake-news detection model. We conduct extensive experiments on the Weibo21 dataset, the results of which indicate that our multi-domain methods perform well, in comparison with baseline methods, on the Weibo21 dataset. Our proposed framework presents greatly improved multi-domain fake-news detection performance.

Keywords: fake news detection; mixture-of-experts model; embedding fusion; social media; pre-trained models

1. Introduction

The traditional mode of information transmission, represented by newspapers and periodicals, has been mostly superseded by online social networks. Through social networks, billions of users around the world connect to the Internet every day and access diverse content, therefore forming a thriving digital society. As such, the Internet has profoundly affected people's lives. Social networking platforms, such as Sina Weibo, Twitter, and so on, have become an important source of news, due to their accessibility and convenience. Although the development of social media provides a convenient information source for people, it has also become a key platform for the spread of fake news. Fake news can have serious and irreversible consequences for individuals and societies. For example, after the 2011 Japan earthquake [1] and Hurricane Sandy in 2012 [2], in September of 2022, the fake news "after typhoon "Meihua" made landfall, Beilun and Cixi in Ningbo, Zhejiang, China were flooded and the reservoir burst" was widely circulated, which aroused widespread societal concern and disrupted public order. Therefore, fake-news detection is not only a technical problem but also an urgent social problem to solve [3]. At present, users can submit any information to social media platforms such as Weibo and Twitter, which may

include fake news. The platform then performs a manual inspection to determine the validity of the reported information. Although this approach can help to limit the spread of fake news, it relies on human review and expert knowledge, while the fake news may have already been widely circulated during the manual review process.

To address this problem, some works [4–8] have focused on fake-news detection in a single domain. Zhou et al. [9] have classified the current approaches for fake-news detection based on knowledge graphs, news genres, distribution patterns, and credibility networks. From a data mining perspective, Shu et al. [10] divided fake-news detection research into feature extraction and model creation. Zubiaga et al. [11] provided an overview of existing research on fake-news detection tasks, including event authenticity detection, event tracking, stance classification, accuracy classification, and other related goals. Numerous additional studies [12–14] have utilized multimodal learning to address the issue of classifying fake news. EANN was proposed by Wang et al. [14], which considers event categorization as an auxiliary task to assist with feature extraction. The event categorization branch enables the extraction of both event-specific and event-aware information by better decoupling the mined multimodal characteristics. Unimodal feature extractors were employed by Dhruv et al. [13] to analyze pictures and text, and multimodal VAEs were then used to extract common representations from them. The decoder then makes an effort to recreate the original text and low-level picture characteristics using the sampled representation provided by the VAE. Other efforts utilize extra information from the dataset in addition to the network design focus. For instance, Qi et al. [15] have suggested manually extracting this type of information as a linguistic aid, as they claimed that image feature extractors cannot adequately recognize visual elements such as celebrities, landmarks, and language inside images. To assess the difference in sentiment between posts and comments, Zhang et al. [16] developed a unique dual sentiment feature descriptor and confirmed that the use of dual sentiment can discriminate between fake and true news.

In some recent studies, Mansour Davoudi et al. [17] proposed a fake-news detection model, which is mainly composed of the following three parts: dynamic analysis, static analysis, and structural analysis. Dynamic analysis encodes evolutionary patterns using recurrent neural networks. In addition, static analysis represents the overall characteristics of the network by making use of fully connected networks. Structural analysis encodes the structure of the network using the node2vec algorithm. Sonal Garg et al. [18] focused on various categories of linguistic features for the effective identification of fake news, including complexity features, readability feature indices, psycholinguistic features, and stylistic features. Alex Munyole Luvembe et al. [19] proposed a mechanism based on depth-normalized attention to enrich and extract dual affective features, and then proposed an adaptive genetic weight update random forest (AGWu-RF) for classification. Gongyao Jiang et al. [20] and proposed knowledge Prompted learning (KPL) for this task. First, we applied just-in-time learning to FND by making complex cueing templates and corresponding spoken words for the task. Second, we incorporate external knowledge into the cueing representations to make them more expressive to predict spoken words. Experimental results on two benchmark datasets show that just-in-time learning outperforms baseline fine-tuned PLM utilization of FND and can outperform all previous representative methods. Furthermore, the final knowledge model (i.e., KPL) can provide further improvements. Compared with traditional media, although detecting fake news with a single domain is useful, contemporary news and postings often contain information in several domains [21], and it is difficult for a single-domain fake-news detection method to deal with the growing volume of fake news. Therefore, methods for the detection of multi-domain fake news can be adapted to current actual usage scenarios.

We propose a new multi-domain fake-news detection framework, which is a mixture of expert networks based on pre-trained and collaboration modules for fake-news detection. Specifically, the text is encoded by a fine-tunable BERT [22] and a CLIP text encoder [23]. BertTokenizer [22] is used to encode the text of the dataset, and the text encoder of CLIP is jointly called for encoding. As the framework of the text encoder is based on the BERT

model, it has high semantic space consistency. The fusion feature E_{fusion} is generated by connecting the features generated by the CLIP branch with the Representation Branch, which avoids the further introduction of noise and redundant features during the feature fusion process. The key component of our framework is the Collaborative Branch, in which the sentiment module is used to analyze the inherent sentimental information of the data. Previous studies have shown that fake news is often accompanied by strong sentiments, which are significantly different from real information [16]. Therefore, it is important to fully exploit the inherent emotional information of fake-news language patterns [24]. We use the attention module to obtain a sentence-level embedding and introduce the inherent domain embedding to form the collaboration module, which can adaptively determine the weights of the expert models to enhance or suppress their contribution in the final mixture-of-experts model. This module is theoretically compatible with most mixture-of-experts models and multimodal guided learning methods. Finally, a mixture-of-experts model based on TextCNN [25] is used for learning, and a classifier is connected to distinguish fake news from real news. We conducted a range of experiments on the Weibo21 dataset, the results of which indicate that the performance of the proposed multi-domain methods is comparable to that of several multi-domain learning baselines [14,21,26–28] on the Weibo21 dataset and several traditional classification fake-news detection methods [8,22,25]. In particular, our model framework obtained significantly improved detection accuracy.

The key contributions of this paper mainly comprise three aspects:

- We propose a novel multi-domain fake-news detection framework; in particular, a mixture-of-experts model-based network based on a pre-trained representation embedding module and a collaborative module for fake-news detection.
- We propose a collaborative module that can adaptively determine the weights of the expert models to enhance or suppress their contributions to the mixture-of-experts model. This module is theoretically compatible with most mixture-of-experts models and multimodal learning methods.
- We conduct extensive experiments on the Weibo21 dataset, and the results indicate that our model framework can achieve significant improvements over the considered baseline methods.

The remainder of this paper is structured as follows. We present the related work in Section 2, including background information and literature reviews on Fake-News Detection, Mixture of Expert Models, and CLIP [23]. Section 3 provides a detailed description of our proposed method, including the Content Embedding and Collaborative Branch, and the mixture-of-experts model. The Content Embedding includes the CLIP branch, for which we introduce the use of CLIP and training models in our framework in detail. Meanwhile, the Representation Branch uses BertTokenizer [22] to obtain the text embedding. The Collaborative Branch includes the Sentiment and collaboration modules. The final mixture-of-experts model details include the definitions of the expert model, the classifier, and the final loss function. Section 4, we first introduce the dataset and baseline methods, and then present the experimental details. We conduct a detailed performance comparison between some traditional classification baseline [8,22,25] and multi-domain baseline [14,21,26–28] models. The performance analysis and assessment of the performance of the expert models are detailed in Section 4.3. We describe the result of ablation experiments in Section 4.4, to verify the effectiveness of each module. Section 5 provides a discussion of the limitations of this study and potential areas for future research. Finally, in Section 6, we summarize our research conclusions.

2. Related Work

2.1. Fake-News Detection

Fake-news detection has always been an important and popular research topic in the domain of artificial intelligence. To solve this problem, early machine learning methods mainly relied on manual feature extraction [4,6,29,30]. However, the selection and design of features is challenging; it is difficult to obtain high-dimensional, complex, and abstract

feature data; and the obtained feature vectors often lack robustness [31]. In contrast, deep learning-based fake-news detection methods [2,32–36] have shown great potential in this domain. Ma et al. [32] proposed the use of recurrent neural networks and their various variants for the detection of fake news. Yu et al. [34] suggested using CNN networks to extract the semantic features of text for fake-news detection. Subsequent studies have added Gated Recurrent Unit (GRU) [37] and Long Short-Term Memory (LSTM) [2] into models to improve the detection effect. User behavior characteristics have also been employed in fake-news detection methods [29,38–44]. For example, Morris et al. [42] proposed a fake-news detection method based on user behavior characteristics. They discovered that, if the number of followers is significantly higher than the number of fans, the user's reputation suffers. Suzuki et al. [40] fused user forwarding behavior features to judge the credibility of tweets. Mohammad et al. [41] proposed features related to the client and release site, and integrated more than a dozen types of features for fake-news detection. Parisa Bazmi et al. [45] proposed a Multi-View Co-Attention Network (MVCAN) that jointly models the potential topic-specific credibility of users and news sources for fake-news detection. The key idea is to represent news articles, users, and news sources by encoding the subject views of news articles, users' SC biases (determining users' perceptions of sharing news), and partisan biases of news sources as vectors. Hu et al. [46] formulated the fake-news detection task as a causal graph reflecting causal factors and accordingly proposed a novel framework–Causal Inference Using Image-Text Matching Bias in Multimodal Fake-News Detection. Shufeng Xiong et al. [47] proposed a two-round inconsistency-based multimodal fusion network (TRIMOON) for fake-news detection, which consists of three main parts: multimodal feature extraction module, feature fusion module, and classification module.

In terms of Chinese datasets, Wu et al. [29] used user forwarding features to detect fake news in 2015, including topics, user information, forwarding time, and text. Gao et al. [44] proposed a fake-news detection method based on publisher features, microblog text features, and dissemination features, which achieved the purpose of judging fake news with high accuracy through feature fusion. Yang et al. [43] utilized an unsupervised method to detect fake news by building a Bayesian classifier and extracting user features, message features, and user opinions. Other types of detection models [14,21,48,49] have also been proposed. Ma et al. [49] and Li et al. [48] integrated related tasks into fake-news detection approaches to improve accuracy. Zou et al. [14] adopted a minimax game approach to extract event-invariant features, but ignored domain-specific features. Yin et al. [21] proposed a combined model that maintains domain-specific and cross-domain knowledge to detect fake news from a multi-domain standpoint, but did not make full use of domain information. Qiong Nan et al. proposed a domain and instance-level transmission framework (DITFEND) for fake-news detection, which improves the performance of specific target domains. Zhu et al. [50] and proposed an entity debiasing framework (ENDEF) that generalizes the fake-news detection model to future data by mitigating entity deviance from a causal perspective.

In the real world, news can be classified into multiple domains, and a given piece of news can belong to one or many of these domains. None of the above-mentioned fake-news detection works have significantly focused on multi-domain fake-news detection.

2.2. Mixture-of-Experts Model

The mixture-of-experts model, which involves jointly learning a set of domains, has been shown to be beneficial in a variety of applications [26–28,51–53]. The idea of this model is to train multiple neural networks, where each neural network (expert) is assigned to a different part of the dataset. Each expert in the system will have an area of data that it is targeted at and, so, performs better than other experts in this domain. The use of a mixture-of-experts system can resolve the problem in which a single model is only good at dealing with a certain kind of data. As the size of the dataset increases, the learning performance of the mixture-of-experts model will be significantly improved.

Ma et al. [26] proposed the multi-door mixture-of-experts (MMoE) model, a revolutionary multi-task learning approach. The sequential mixture-of-experts (MoSE) model, introduced by Qin et al. [27], specifically models sequential user behavior by utilizing Long Short-Term Memory (LSTM). Zhu et al. [53] proposed a new methodology with two alignment stages that align the distribution of each source–target domain pair in numerous specialized feature spaces. These studies have concentrated on capturing the inter-task relationships and various representations, as changes in the relationships between tasks can reinforce each task.

2.3. CLIP

CLIP [23] is a pre-training model, which stands for Contrastive Language-Image Pre-Training. It has become a classic model in the domain of multimodal research in recent years. It uses contrastive learning on 400 million image–text paired datasets collected from the Internet to obtain an effective and scalable pre-trained model with strong zero-shot capabilities.

CLIP has two modes: A text mode and a visual mode, corresponding to the text encoder and image encoder, respectively. The text encoder is used to encode text to obtain its embedding, while the image encoder is used to encode pictures to obtain their embeddings. Both embeddings are single vectors of a certain length. The text encoder is a basic BERT model [22], which is essentially composed of self-attention modules; therefore, the structure of the text encoder and the image encoder is basically the same. In CLIP, the text encoder is composed of 12 transformer encoder layers. As text information is simpler than visual information, the text encoder used by each scale of CLIP does not change, and the size remains the same. We observed that little research in the field of fake-news detection has taken the state-of-the-art in multimodal learning into account, which inspired us to apply CLIP-based pre-training to further enhance the performance of the proposed model. In our framework, we only use the text encoder part of the trained CLIP.

3. Approach

In this section, we present a novel framework for multi-domain fake-news detection that uses standard methods for binary classification tasks. The overall architecture of our proposed framework is illustrated in Figure 1. Our framework comprises a CLIP branch, a representation extraction branch, a collaborative branch, and a mixture-of-experts model module. We provide detailed introductions to the various branches and modules in the following sections.
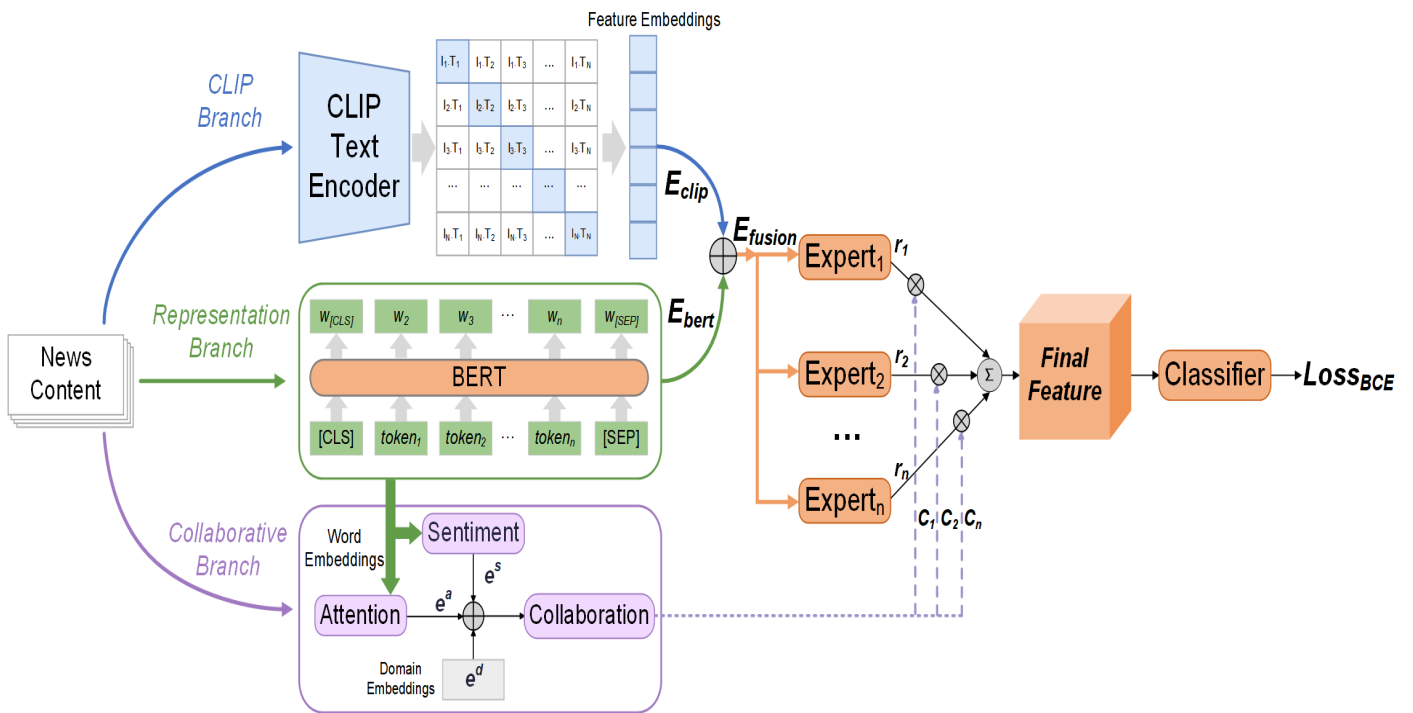
Figure 1. Overview of the proposed framework. CLIP branch consists of a pre-trained CLIP text encoder to obtain embedded news content E_{clip} as shown in Equation (1). In addition, the Representation Branch is composed of a pre-trained BERT to obtain the embedded E_{bert} of news content, which is shown in Equation (2). The fusion embedding E_{fusion} is made of a combination of E_{clip} and E_{bert}, as shown in Equation (4), and then it is input into the expert mixture model for training. The Collaborative Branch includes sentence-level embedding e^a from attention, emotional embedding e^a from Bert's emotion classification model, and domain embedding e^a including various domain features. When these embeddings are input into the collaboration module, the weights of the expert model will be adaptively determined, as shown in Equation (3). Finally, a mixture of expert models is used to study features to achieve a high-performance fake-news detection model, such as Equations (5)–(8).

3.1. Content Embedding

CLIP branch. CLIP [23] has a wide range of applications. A key benefit of using CLIP is that it can handle various types of datasets, as it can employ any type of text description, including labeled and unlabeled text. Additionally, because CLIP is an end-to-end model, the issue of manually designing features which is present in many conventional techniques can be avoided. Notably, text features play a leading role in fake-news detection.

We first use the CLIP text encoder [22] in the CLIP branch to encode the text content into an embedding that is suited to the trained model. We use the 'ViT-B/32' pre-trained CLIP model and freeze the weights. The input of the model is the news content of the dataset text, and the text encoder is used to obtain the embedding. These vectors are mapped into a joint multi-channel space, which is essentially a feature matrix. There are n positive samples on the diagonal of the feature matrix, while the $n^2 - n$ elements on the off-diagonal of the matrix are negative samples. We obtain new vectors I_n and T_n in the existing feature matrix, and the feature embedding E_{clip} is obtained as the output, whose maximum value is $I_i \cdot T_i = MAX(I_1 \cdot T_1, I_2 \cdot T_2, \cdots, I_n \cdot T_n)$ $(1 \leq i \leq n)$.

$$E_{clip} = MAX(I_1 \cdot T_1, I_2 \cdot T_2, \cdots, I_n \cdot T_n), \qquad (1)$$

Representation Branch. The pre-trained BertTokenizer [22] model has a strong embedding representation capability, and the feature matrix of the BERT-based Tokenizer can perform downstream tasks. To process news articles, we first tokenize the content using BertTokenizer [22] in the Representation Branch, which needs to be encoded into a code that

the model can recognize. Special classification tokens (i.e., [CLS]) and separation tokens (i.e., [SEP]) are added, producing a list of tokens in the shape of [[CLS], $token_1$, ..., $token_n$, [SEP]], where n is the number of tokens in the news article. We then input these tokens into BertTokenizer to obtain the word embeddings W, where $W = [w_{[CLS]}, w_1, ..., w_n, w_{[SEP]}]$. We ultimately obtain the output E_{bert}. During this process, all word embeddings are processed through a masked attention network to obtain sentence-level embeddings.

$$E_{bert} = [w_{[CLS]}, w_1, ..., w_n, w_{[SEP]}], \tag{2}$$

Additionally, we input the embeddings into the sentiment and attention modules. The sentiment module obtains embeddings for sentiment categorization. To capture the personalized representation of each sentence, we define a learnable sentiment vector e_s to assist the experts. The sentiment module acquires sentiment embeddings, which provide the Collaborative Branch with specific sentiment information inherent to fake news, therefore enriching the fine-grained personalized representation. The attention module consists of a mask-attention network to obtain the sentence-level embedding e_a.

3.2. Collaborative Branch

Sentiment Module. Sentiment analysis is the process of analyzing text to determine whether a message has a positive, negative, or neutral sentimental tone, utilizing a sentiment module to provide objective insights. Fake news is often associated with strong sentiments, significantly different from real information [16]. To leverage these inherent sentiment patterns in fake news, we designed a sentiment analysis module that accurately detects sentimental information in news texts to assist in identifying fake news. For the sentiment analysis module, we used the Weibo_senti_100k [54] dataset to fine-tune the pre-trained BERT model [22].

To capture the personalized representation of each sentence, we define a learnable sentiment vector e_s. This module utilizes sentence embeddings to guide the expert models and outputs a vector e^s. The sentiment analysis module is denoted as $S(\cdot; \beta)$, where β is a parameter of the module. Finally, the sentimental feature vector of the news is obtained through this module.

Collaboration Module. When creating high-quality representations of news across multiple domains using traditional methods, simply averaging the representations of all experts may result in the loss of domain-specific information. Therefore, we propose the use of a collaborative module to adaptively predict the influence of experts in the mixture-of-experts model, to enhance or suppress the contribution of each in the final model. The core of our framework consists of a sentiment module and an attention module, as well as a feed-forward network-collaboration module, which determines the degree of contribution from each collaborator (i.e., expert). Our collaborative framework can theoretically be integrated into any existing mixture-of-experts model or multimodal learning task.

The embedding we obtained from the Representation Branch is obtained through the attention module to obtain the sentence-level embedding e^a, the embedding e^s obtained by the sentiment module, and the domain embedding e^d obtained by the embed every domain whole-content, which are input into the feed-forward network-collaboration module through addition and fusion of the embeddings. The collaboration module consists of linear layers, ReLU layers, and a SoftMax classification function. Therefore, the Collaborative Branch outputs a vector C_i $(0 \leq i \leq n)$ that guides the weight ratio in the mixture-of-experts model. The collaboration module is denoted as $C(\cdot; \beta)$, where β is the parameter of the collaboration module, n is the number of experts, and i is the particular identifier of an expert.

$$C_i = softmax\left(C\left(e^a \oplus e^d \oplus e^s; \beta\right)\right), \tag{3}$$

where the feed-forward network used for the collaboration module is denoted as $G(\cdot; \beta)$, and the input to the collaboration module consists of the domain embedding e^d, the

sentence embedding e^s, and sentiment embedding e^s. The output of the collaboration module is normalized to $G(\cdot)$ using a SoftMax function, and the resulting weight vector $G_i \in \mathbb{R}^n$ represents the relative importance of each expert model in the final representation, as shown in Equation (3).

3.3. Mixture-of-Experts Model

We use numerous expert networks to extract various news representations; in particular, we utilize TextCNNs [25]. This work proves a simple CNN with little hyperparameter tuning and static vectors achieves excellent results on multiple benchmarks. Each expert network is denoted as $\lambda\left(E_{fusion}; \theta_i\right) (1 \leq i \leq n)$:

$$E_{fusion} = E_{clip} \oplus E_{bert}, \tag{4}$$

$$r_i = \lambda\left(E_{fusion}; \theta_i\right), \tag{5}$$

where E_{fusion} consists of the BERT embeddings plus the CLIP embeddings, θ_i represents the set of parameters to be learned, and n represents the total number of expert networks. Each expert network's output r_i, represents the representation extracted by the corresponding expert network. Using multiple expert networks is advantageous, as a single expert network is only good at extracting representations for one domain. As such, a single expert network may extract only partial information about the news content, which cannot fully cover the characteristics of the news article. Additionally, some fake news articles may not belong to a single domain, such that separate classification of the dataset into a single domain or with limited expert models may affect the detection accuracy.

Classifier and loss function. The final network for fake-news detection is designed as an MLP. First, the output r_i of all expert networks is accumulated and multiplied by the collaborative influence function C_i output by the collaboration module:

$$Final\ Feature = \sum_{i=0}^{n} C_i \otimes r_i, \tag{6}$$

We feed the news feature vector S into the MLP, then perform classification at the SoftMax output layer:

$$\hat{y} = softmax(MLP(Final\ Feature)), \tag{7}$$

Finally, we train the binary classifier to categorize news as fake or not using the binary cross-entropy loss function (BCELoss). The actual and predicted labels for the ith news sample are designated as y^i and $\hat{y}^i$, respectively:

$$Loss_{BCE} = -\sum_{i=1}^{n}\left(y^i log\hat{y}^i + \left(1 - y^i\right) log\left(1 - \hat{y}^i\right)\right). \tag{8}$$

4. Experiments

In this section, we provide a detailed description of the dataset used in our experiments. We also introduce the baseline model, evaluation metrics, and parameter settings. Furthermore, we also analyze and compare the efficacy of our proposed multi-domain fake-news detection framework with other baseline approaches on the dataset. Finally, we also conduct ablation tests to validate our design decisions.

4.1. Experimental Setup

DataSets. The Weibo21 dataset [28] was used to evaluate the performance of our proposed multi-domain fake-news detection technique. This dataset was gathered from the Sina Weibo [55] platform between December 2014 and March 2021, and contains 9128 news

data instances. This dataset contains news text, image content, timestamps, and comments for each data point, which is tagged as either true or fake news. For the fake data, the news clips were officially judged as fake by the Weibo community management center [56]. For the real data instances, the real news clips contemporaneous with the fake news clips were verified using NewsVerify [57], a tool dedicated to detecting and confirming fraudulent news snippets on Weibo. Reference is made to news classifications from several well-known fact-checking websites, as well as some research papers and reports, including Vosoughi et al. [58]. "2017 Tencent Rumor Control Report" and China Internet Joint Rumor Refuting Platform. The dataset is divided into nine categories: science, military, education, disaster, politics, health, finance, entertainment, and society. The statistics of the collected dataset are provided in Table 1 and the examples data as shown in Table 2. To ensure the impartiality of labeling, ten experts were hired to manually label the news. Eventually, 4488 pieces of fake news and 4640 pieces of real news were labeled.

Table 1. Weibo21 dataset statistics [28].

Domain	Science	Military	Education	Accidents	Politics	Health	Finance	Entertainment	Society	All
Real	143	121	243	185	306	485	959	1000	1198	4640
Fake	93	222	248	591	546	515	362	440	1471	4488
All	236	343	491	776	852	1000	1321	1440	2669	9128

Table 2. Examples data of Weibo21 dataset.

Content	Domain	Fake Label
【熊猫宝宝地震了也会找警察】雅安是大熊猫栖息地...警察叔叔的腿。	Accidents	0
今晚有三首歌是张杰以前唱过的，不同的声音...回味一下杰哥的版本吧。	Entertainment	0
宝宝夏天不能吹空调，吹了就会得空调病?	Health	1
在过去，要修建一座堡垒，需要花费好几个月...里面的设施应有尽有。	Military	1
每天早上6点20，武昌工学院某群便炸开...发红包方式叫学生起床。	Education	0

Baseline Methods. To validate the performance of our proposed fake news detection framework, we compared it with several baseline methods, including traditional classification models [8,22,25], and multi-domain [14,26–28] baseline models.

Traditional classification baseline methods are considered to have achieved more improvements and can perform robustly in various learning tasks [59]. For comparison with traditional classification models, we selected TextCNN [25], BiGRU [8], and BERT [22]. TextCNN [25] uses a simple CNN network and has achieved excellent results on multiple benchmarks. BiGRUe [22] has been proposed to detect fake news faster and more accurately based on the use of an RNN. BERT [22], based on the Transformer architecture[60], utilizes a bidirectional encoder representation and has obtained state-of-the-art results in 80 natural language processing tasks. The EANN [14] model uses only the text branch to extract domain-independent features. MMOE [26] and MOSE [27] have been designed for multi-task learning. EDDFN [21] models different domains to preserve domain-specific and cross-domain knowledge. MDFEND [28] is a model designed for the detection of multi-domain fake news.

In the traditional classification baseline, we experimented with one model at a time across all domains, calculated the f1-scores separately for each domain, and computed the final column using data from all domains. The input features of TextCNN [25] were embedded by Word2Vec [61], and the convolutional structure was modified to be the same as that of the expert modules. BiGRU [22] inputs sequentially for each news item, therefore preserving the sequential information.

4.2. Experimental Details

Parameter Settings. All our experiments were performed on a server terminal with Ubuntu 18.04.5 LTS, Intel Xeon 6126 2.60 GHz CPU, and $4 \times$ V100 GPU.

We used the same parameters for all the approaches, to ensure a fair comparison. In all models, the MLP had the same structure, with one dense layer having 384 hidden units. The length of the text in this dataset was mostly distributed around 0–100. About 90% of the microblog texts were less than 170 characters. If the input length of the model was too large, the vector may be sparse, affecting the performance of the model. Therefore, the maximum length of the input sentence was set to 170 words; that is, the max_seq_length parameter for the BERT model was 170. When the text length exceeds this value, it will be truncated; if it is insufficient, it will be filled with zeros, such that the input text length remained aligned. The embedding vector dimension of a word was fixed at 768 for BERT [22] and 200 for Word2Vec [61,62]. We utilized the Adam [63] optimizer to determine the best learning rate between 1×10^{-6} and 1×10^{-2}. All methods used a batch size of 64. We repeated this technique ten times, to strengthen the credibility of our experiments.

CLIP branch. We used the translate tool API [64] to translate Chinese text from Chinese to English, as CLIP [23] does not have a pre-trained Chinese text model. Additionally, to comply with the requirement that there is a maximum size for text input in CLIP, we constructed summary sentences for texts longer than 50 words using a summary generation model [65]. We employed the 'ViT-B/32' pre-trained CLIP model with frozen weights.

Sentiment Module. The sentiment module was trained by inputting the Weibo_senti_100k dataset into BERT [22]. The Weibo_senti_100k dataset contains about 100,000 Sina Weibo articles with sentimental annotations in the real environment, and about 50,000 positive and negative comments, respectively. Due to the use of real Weibo text in the dataset, a substantial amount of noise data that lacks practical value in sentiment analysis is present. Furthermore, the inclusion of such data increases the data dimension, which can reduce the quality of the analysis results. Therefore, data cleaning was necessary to eliminate noise and irrelevant information not related to sentiment expression. For this, we utilized various data cleaning procedures, including the removal of digits, stop words, URL links, "@" signs, and extraneous punctuation marks. In the end, we used the cleaned dataset to build a sentiment module.

The embedding vector dimension of a word was fixed at 768 for BERT [22]. We utilized the Adam [63] optimizer to obtain the best learning rate between 1×10^{-6} and 1×10^{-2}. The batch size was 64 and we used the binary cross-entropy loss function (BCELoss).

Evaluation Metrics. The average f1-score was used to assess the overall performance of all models, which were tested for TP, TN, FP, and FN, which stand for true-positive, true-negative, false-positive, and false-negative counts, respectively. The precision (P) was calculated using Equation (9), recall (R) using Equation (10), and f1-score ($F1$) using Equation (11), as follows:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{11}$$

4.3. Performance Comparison

Performance Comparison. Table 3 shows the performance comparison our proposed framework and the other comparative frameworks on the Weibo21 dataset. The f1-scores for different domains (e.g., Science, Military, Education, and so on) are provided, and the ALL column shows the average f1-score across all domains. The results for the proposed model were improved in all domains except the Military domain. The results indicate that it is a reliable and robust fake-news detection algorithm, which can detect and classify fake news in multiple domains. Figures 2 and 3 provide a visual comparison of performance in various

domains. And Figure 4 shows the performance comparison on our proposed framework and the other comparative frameworks on R(recall), P(precision), AUC, F1(f1-score).

We further compared our framework with the aforementioned cutting-edge techniques and obtained the following conclusions from the experimental data. First, the results of traditional models were not as good as those of multi-domain models, indicating the complexity of news domains in practical fake-news detection settings and the importance of multi-domain fake-news detection research. However, in the finance domain, the BERT traditional model outperformed the most multi-domain model, indicating that traditional models can perform better than multi-domain models in certain domains. Third, the multi-domain models performed better than traditional classification models, highlighting the strength of multi-domain learning in fake-news detection. Furthermore, these results suggest that combining fake-news data from different domains may have negative effects, therefore demonstrating the importance of the adaptive collaboration model module, which affects the influence of the adaptive expert models. Sentimental analysis provides the inherent sentimental information related to domains and sentences, to better model the relationship between domains and achieve higher detection accuracy. Overall, our model outperformed the other multi-domain models.
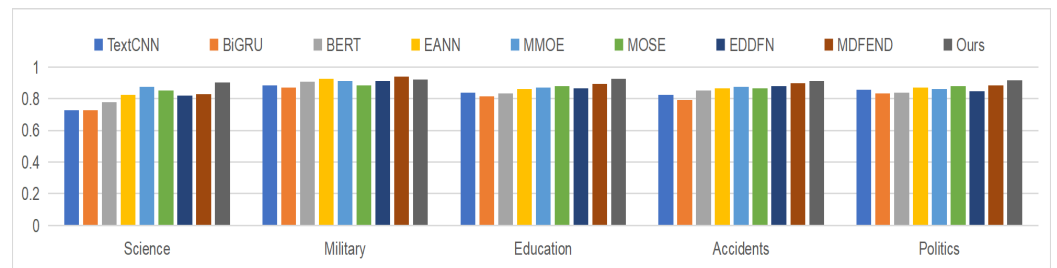

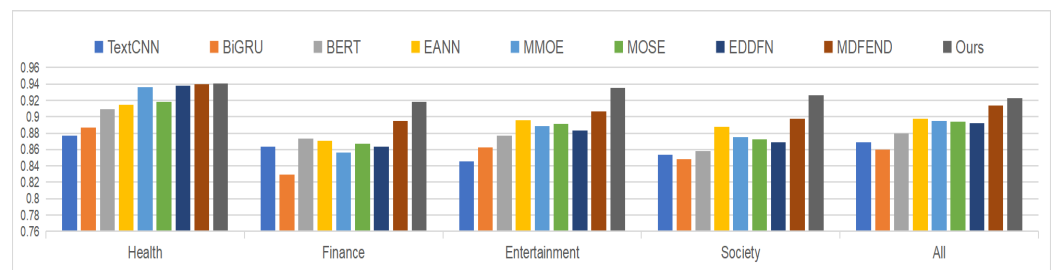**Figure 2.** Multi-domain fake-news detection performance (a).

Figure 3. Multi-domain fake-news detection performance (b).

Table 3. Multi-domain fake-news detection performance (f1-score) on Weibo21 dataset.

Model	Science	Military	Education	Accidents	Politics	Health	Finance	Entertainment	Society	All
TextCNN	0.7254	0.8839	0.8362	0.8222	0.8561	0.8768	0.8638	0.8456	0.8540	0.8686
BiGRU	0.7269	0.8724	0.8138	0.7935	0.8356	0.8868	0.8291	0.8629	0.8485	0.8595
BERT	0.7777	0.9072	0.8331	0.8512	0.8366	0.9090	0.8735	0.8769	0.8577	0.8795
EANN	0.8225	0.9274	0.8624	0.8666	0.8705	0.9150	0.8710	0.8957	0.8877	0.8975
MMOE	0.8755	0.9112	0.8706	0.877	0.8620	0.9364	0.8567	0.8886	0.8750	0.8947
MOSE	0.8502	0.8858	0.8815	0.8672	0.8808	0.9179	0.8672	0.8913	0.8729	0.8939
EDDFN	0.8186	0.9137	0.8676	0.8786	0.8478	0.9379	0.8636	0.8832	0.8689	0.8919
MDFEND	0.8301	0.9389	0.8917	0.9003	0.8865	0.9400	0.8951	0.9066	0.8980	0.9137
Ours	0.9049	0.9204	0.9263	0.9109	0.9169	0.9407	0.9184	0.9353	0.9266	0.9223

Performance Analysis. Our framework generally outperformed the state-of-the-art methods for the following reasons. First, the use of the pre-trained CLIP encoder in the framework allows for the generation of text features with rich semantic information in the same semantic space; especially considering its powerful zero-shot ability, which greatly

helps the framework when a single text belongs to a multi-domain. It also provides supplementary information for the mixture-of-experts model. The Collaborative Branch allows for adaptive determination of the weights for the expert models, to enhance or suppress their effects, therefore avoiding the impact of invalid features on the final feature representation ability and further improving the classification accuracy.

Many multi-domain fake-news detection methods, such as EANN [14], obtain fusion features directly through linking or attention mechanisms. However, these features alone are not enough to distinguish fake news, as the extracted semantic features and image information fusion method are not suitable for multi-domain classification. For example, EDDFN [21], a cross-domain fake-news detection model, can represent different domains. The sampling method for its LSH is used to avoid unseen problems. But there is no unseen problem in the training set. Therefore, the experimental results of these methods are unsatisfactory.
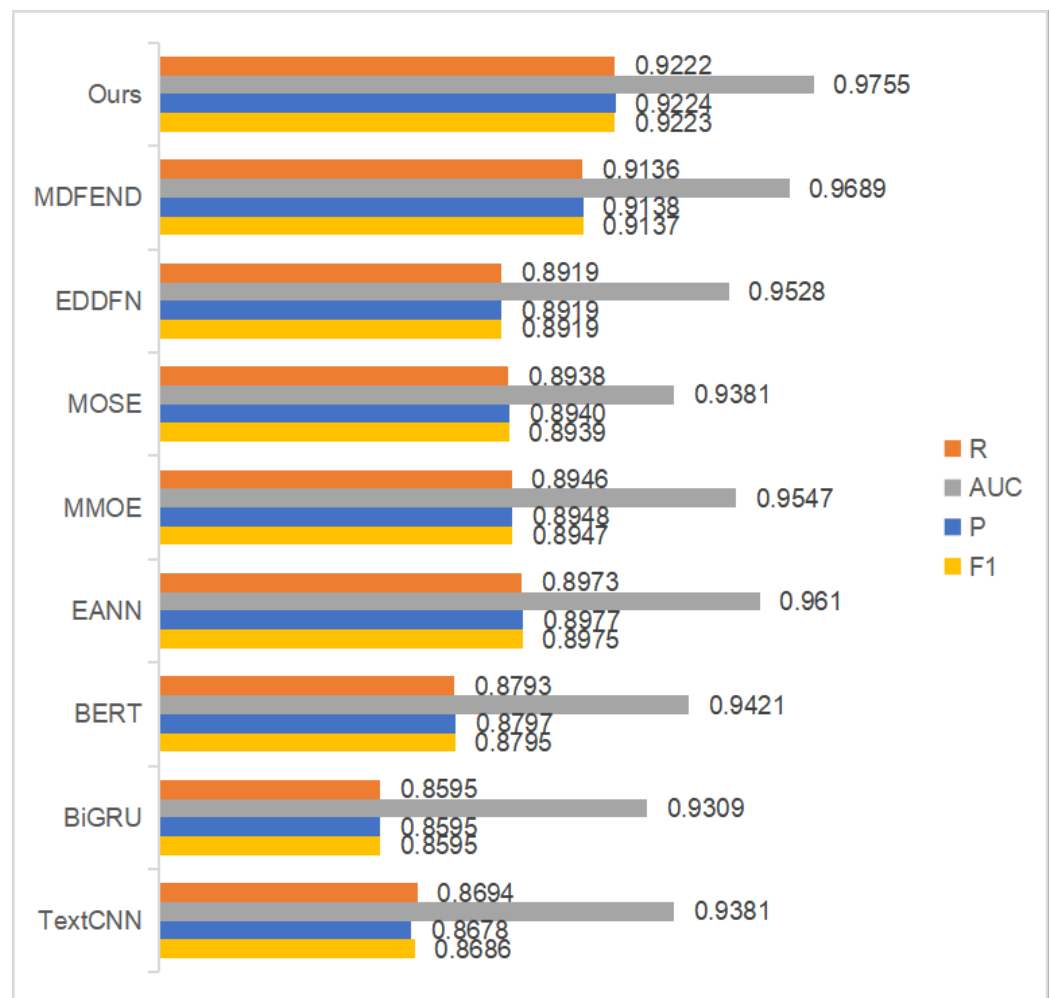

Figure 4. Performance comparison of R(recall), P(precision), AUC, F1(f1-score).

Performance of Experts. Figure 5 illustrates the impact of different numbers of expert models on the final predicted classification results. When the number of experts was set to 90, the lowest score of 0.88 was achieved; meanwhile, when the number of experts was 50, the highest score of 0.91 was achieved. It can be inferred that the choice of the number of experts in the mixture-of-experts model has a great impact on the final performance results. Our proposed collaboration module can adaptively predict the individual influence of expert models to enhance or suppress the contributions of different expert models to the prediction results, therefore avoiding the influence of too much contribution by unfavorable

experts in the text domain and giving full play to the contributions of favorable experts. This module led to improved performance, as shown in column CM of the Figure 5.
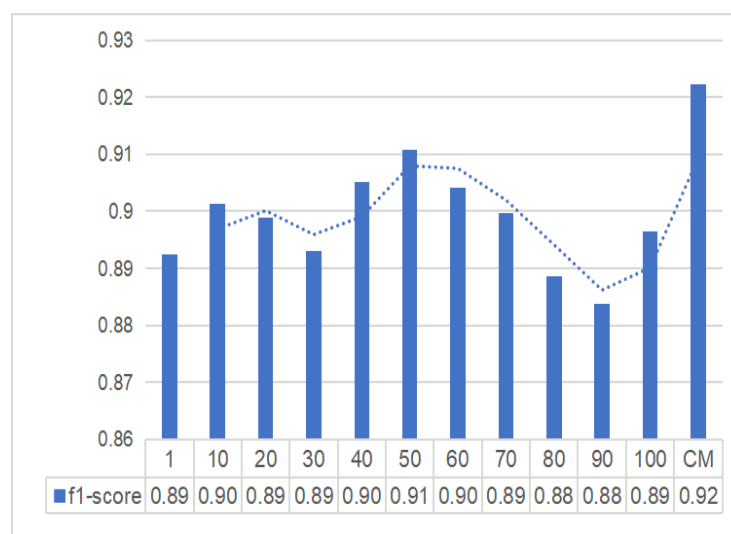

Figure 5. The effect of different numbers of expert models on performance (based on f1-score). CM, collaboration module.

4.4. Ablation Study

We conducted a detailed ablation study on our proposed framework and collaboration module by combining different key components, to evaluate the performance of different parts of the model and to validate the design and choice of our algorithm. The results are presented in Table 4 and Figure 6. In each test, we removed different components.

Our proposed framework. The first line is without the CLIP Branch; that is, only the Representation Branch and Collaborative Branch were retained. It can be seen that CLIP encoding can provide a lot of useful information to assist the mixture-of-experts model, but the main performance is provided by the embedding provided by the representation part, which provides effective information for the classifier. It can effectively provide discriminative features for fake-news detection tasks and significantly improve classification accuracy. Its powerful zero-shot capability greatly helps the framework to judge text classification and achieve higher scores.

The second line is without the Collaborative Branch; that is, only the CLIP Branch and Representation Branch were retained. The results indicate that the Collaborative Branch module makes full use of domain embedding and text sentiment information to adaptively determine the weights of expert models to enhance or suppress their effects, therefore avoiding the influence of invalid features on the final feature representation ability and further improving the classification accuracy.

The third line shows a pared-down version of our model, without the CLIP Branch and Collaborative Branch, which only utilizes embeddings from the Representation Branch as input to the mixture-of-experts model for learning. The results demonstrate the powerful representation ability of BERT and the learning ability of the mixture-of-experts model.

Finally, the full version of our framework is detailed on the fourth line, which contains the CLIP Branch, Representation Branch, Collaborative Branch, and the final mixture-of-experts model.

Table 4. Ablation study of the proposed architecture design. The complete architecture received the highest score, showing that every module and modality in the architecture of our method is employed efficiently. First line is without CLIP Branch; second line is without Collaborative Branch; third line is without CLIP Branch and Collaborative Branch. (w/o), without.

Model	Science	Military	Education	Accidents	Politics	Health	Finance	Entertainment	Society	All
(w/o) CLIP	0.8649	0.9015	0.9183	0.8879	0.9166	0.9247	0.9226	0.9110	0.9216	0.9077
(w/o) Collaborative	0.9032	0.9134	0.9435	0.8984	0.9167	0.9233	0.8870	0.9054	0.9264	0.9130
(w/o) CLIP & Collaborative	0.8365	0.9051	0.9181	0.8882	0.8935	0.9215	0.8967	0.9113	0.9074	0.8976
Complete Model	0.9049	0.9204	0.9263	0.9109	0.9169	0.9407	0.9184	0.9353	0.9266	0.9223

Collaborative branch. As shown in Figure 6, the first column, there is without sentiment module, only e^a from the attention module and e^d from domain embeddings are retained. It showed that the sentiment module can provide useful sentiment features to help the Collaborative Branch to assist the mixture-of-experts model better.

The second column does not have domain embeddings, only e^a from the attention module and e^s from the sentiment module are retained. It can be seen that domain embeddings are helpful to the performance of the final collaboration module. It can provide the inherent domain feature of each domain.

In the third column, there is without sentiment module and domain embeddings, only the attention module is retained. It showed that the basic collaboration module is helpful for the fake-news classification of the mixture-of-experts model.

The fourth column contains the whole module of the Collaborative Branch, e^s from the sentiment module, e^d from the domain embeddings and e^a from the attention module. The chart shows that the Collaborative Branch of our framework is an optimistic impact on the mixture-of-experts model's performance.
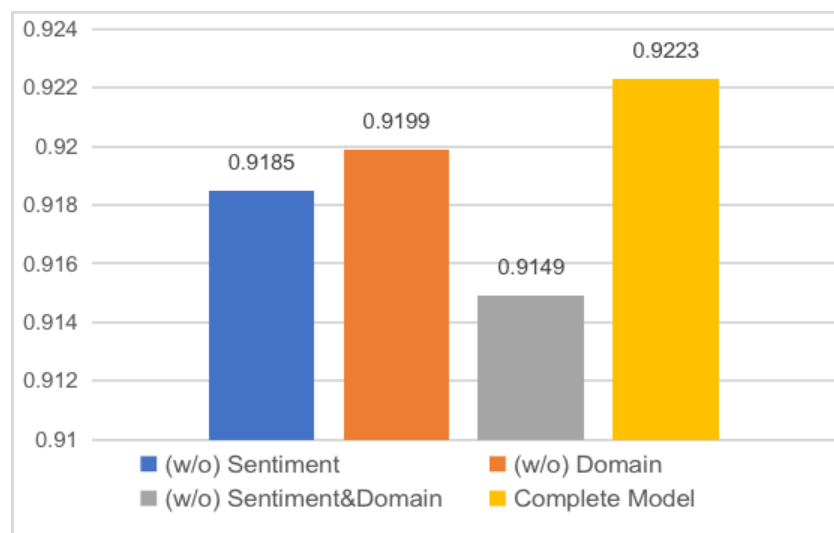

Figure 6. Ablation study of the proposed architecture Collaborative Branch. The complete architecture received the highest score, showing that every module and modality in the architecture of our method is employed efficiently. The first row is without a sentiment module; the second row is without a domain embedding; the third row is without a sentiment module and domain embedding. (w/o), without.

5. Discussion and Future Work

Limitations. There are several possible reasons for the small improvement in the accuracy of fake-news detection algorithms in some domains: there are serious data imbalances in fake-news data in different areas: for example, in the accident domain and the opposite financial domain, fake news is more common than real news, resulting in poor performance of algorithms in detecting real news. Such imbalance would make the algorithm

more inclined to make predictions for "common categories", while performing poorly on rare categories. (2) Difficulty in feature extraction in specific domains: For example, fake news in the health domain may involve medical terminology and expertise, while fake news in the entertainment domain may be related to celebrities. If the algorithm cannot effectively learn appropriate features from it, it will lead to difficulties in performance growth. (3) Real environment data: The performance improvement of fake-news detection algorithms largely depends on the quality of the data used. Weibo21 is a dataset of real social environments containing a large amount of noise. Mislabeled or low-quality samples, algorithms may learn inaccurate or invalid patterns. Therefore, it is difficult to further improve its performance. For example, military data contains a large amount of news content that should not be part of the military field. (4) Situations and subjectivity: Some fake news may involve context and subjective judgment rather than mere factual statements. For such fake news, improvements in accuracy may be limited by the ability of algorithms to understand context and cause.

Vulnerability. Deep learning models are generally vulnerable to confrontational attacks. This means that malicious attackers can deceive the model by deliberately creating small perturbations, causing the model to make erroneous predictions. The study of confrontational attacks is important to ensure the robustness of the model. The pre-trained model can resist traditional confrontation attacks, but for targeted violence attacks, it may deceive the model to distinguish fake news from real news.

Other social media and foreign language datasets. Theoretically, our framework can be adapted to any social media platform or dataset in different languages, as the proposed framework is not influenced by the language family.

Future work. Despite significant progress in the field of fake-news detection, multimedia social data analysis remains a challenging task with several opportunities for further advancement. At present, fake-news detection research is generally focused on enhancing the quality of single-modal data representations and obtaining higher-quality multimodal fusion features. In the domain of fake-news detection, there are several trends for future development:

1. Improving the quality of original data: The accuracy of subsequent detection and analysis tasks depends on the quality of the original data. However, most original data suffers from issues such as incompleteness, sparsity, and imbalance. Therefore, one of the key challenges in future research will be to address the imbalance and poor integrity of the original data.

2. Increasing the diversity of multimodal data: Social multimedia data types include various forms of media, such as social links and location information. The diversity of multimodal data can be increased further by leveraging social media attribute information such as labels, location, and time. Therefore, how to mine more external knowledge should be explored in future research.

3. Integrating information from multiple platforms: Existing research has focused on a single social network, such as only using Weibo posts for fake-news detection without incorporating information provided by WeChat users. As information missing from one platform may be available on others, a thorough synthesis of information from multiple social networks can provide more comprehensive real-world social data. Therefore, the next stage could focus on cross-platform information fusion approaches, such as transfer learning, which can transfer knowledge from one social platform to another.

4. Addressing redundancy and noise in social media data: The growth rate of computer hardware cannot keep pace with the increasing demand for multimedia data. The redundancy of large-scale and ultra-large-scale social media data cannot be ignored while utilizing large-scale multimedia data. To improve data quality while reducing computational efforts, a well-designed data filtering technique may be used.

6. Conclusions

The detection of multi-domain fake news is an important line of research. In this study, we proposed a multi-domain fake-news detection framework based on the mixture-of-experts model. Specifically, the input text was encoded using BertTokenizer while jointly invoking the pre-trained CLIP text encoder, and fusion features were obtained by adding the resulting features together. This avoids introducing noise and redundant features in the process of feature fusion. We also proposed a collaboration module, in which a sentiment module is used to analyze the inherent sentimental information of the text, therefore complementing the inherent sentiments of the news language model. In addition, sentence-level and domain embeddings combined with sentimental embeddings and a feed-forward network forms the collaboration module, which can adaptively determine the weights of the expert models. Finally, the mixture-of-experts model composed of TextCNN is used for learning, and a classifier is connected to achieve high performance in distinguishing fake news in a multi-domain context. We conducted extensive experiments on the Weibo21 dataset, the results of which indicate that the proposed multi-domain fake-news detection framework performs well in comparison with baseline methods on the Weibo21 dataset. In particular, our model framework presented greatly improved multi-domain fake-news detection performance.

In the future, we will extend this framework to multimodal learning to detect fake news. And there are some potential application directions of our work: malicious behavior detection, public opinion monitoring, academic paper identification, medical and health information review, and other emerging technical directions.

Author Contributions: Conceptualization, Z.Z.; Methodology, Z.Z.; Software, Z.Z. and Y.L.; Resources, L.S.; Data curation, L.S.; Writing—original draft, Z.Z.; Writing—review and editing, J.Z. and Z.K.; Visualization, Z.Z.; Project administration, J.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Takayasu, M.; Sato, K.; Sano, Y.; Yamada, K.; Miura, W.; Takayasu, H. Rumor diffusion and convergence during the 3.11 earthquake: A Twitter case study. *PLoS ONE* **2015**, *10*, e0121443. [CrossRef] [PubMed]
2. Gupta, A.; Lamba, H.; Kumaraguru, P.; Joshi, A. Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 729–736.
3. Pennycook, G.; Epstein, Z.; Mosleh, M.; Arechar, A.A.; Eckles, D.; Rand, D.G. Shifting attention to accuracy can reduce misinformation online. *Nature* **2021**, *592*, 590–595. [CrossRef]
4. Castillo, C.; Mendoza, M.; Poblete, B. Information credibility on twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 675–684.
5. Jin, Z.; Cao, J.; Guo, H.; Zhang, Y.; Wang, Y.; Luo, J. Detection and analysis of 2016 us presidential election related rumors on twitter. In Proceedings of the Social, Cultural, and Behavioral Modeling: 10th International Conference, SBP-BRiMS 2017, Washington, DC, USA, 5–8 July 2017; Proceedings 10; Springer: Berlin/Heidelberg, Germany, 2017; pp. 14–24.
6. Kwon, S.; Cha, M.; Jung, K.; Chen, W.; Wang, Y. Prominent features of rumor propagation in online social media. In Proceedings of the 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, USA, 7–10 December 2013; pp. 1103–1108.
7. Ma, B.; Lin, D.; Cao, D. Content representation for microblog rumor detection. In Proceedings of the Advances in Computational Intelligence Systems: Contributions Presented at the 16th UK Workshop on Computational Intelligence, Lancaster, UK, 7–9 September 2016; Springer: Berlin/Heidelberg, Germany, 2017; pp. 245–251.

8.　Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.F.; Cha, M. Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), New York, NY, USA, 6–8 July 2016.

9.　Zhou, X.; Zafarani, R. Fake news: A survey of research, detection methods, and opportunities. *arXiv* **2018**, arXiv:1812.00315.

10.　Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 22–36. [CrossRef]

11.　Zubiaga, A.; Aker, A.; Bontcheva, K.; Liakata, M.; Procter, R. Detection and resolution of rumours in social media: A survey. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–36. [CrossRef]

12.　Chen, Y.; Li, D.; Zhang, P.; Sui, J.; Lv, Q.; Tun, L.; Shang, L. Cross-modal Ambiguity Learning for Multimodal Fake News Detection. In Proceedings of the ACM Web Conference 2022, Lyon, France, 25–29 April 2022.

13.　Khattar, D.; Goud, J.S.; Gupta, M.; Varma, V. MVAE: Multimodal Variational Autoencoder for Fake News Detection. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019.

14.　Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; Gao, J. 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, London, UK, 19–23 August 2018; pp. 849–857.

15.　Qi, P.; Cao, J.; Li, X.; Liu, H.; Sheng, Q.; Mi, X.; He, Q.; Lv, Y.; Guo, C.; Yu, Y. 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, Virtual Event, 20–24 October 2021.

16.　Zhang, X.; Cao, J.; Li, X.; Sheng, Q.; Zhong, L.; Shu, K. Mining dual emotion for fake news detection. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 12–23 April 2021; pp. 3465–3476.

17.　Davoudi, M.; Moosavi, M.R.; Sadreddini, M.H. DSS: A hybrid deep model for fake news detection using propagation tree and stance network. *Expert Syst. Appl.* **2022**, *198*, 116635. [CrossRef]

18.　Garg, S.; Kumar Sharma, D. Linguistic features based framework for automatic fake news detection. *Comput. Ind. Eng.* **2022**, *172*, 108432. [CrossRef]

19.　Luvembe, A.M.; Li, W.; Li, S.; Liu, F.; Xu, G. Dual emotion based fake news detection: A deep attention-weight update approach. *Inf. Process. Manag.* **2023**, *60*, 103354. [CrossRef]

20.　Jiang, G.; Liu, S.; Zhao, Y.; Sun, Y.; Zhang, M. Fake news detection via knowledgeable prompt learning. *Inf. Process. Manag.* **2022**, *59*, 103029. [CrossRef]

21.　Silva, A.; Luo, L.; Karunasekera, S.; Leckie, C. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 557–565.

22.　Kenton, J.D.M.W.C.; Toutanova, L.K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL-HLT, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.

23.　Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning PMLR, Virtual, 18–24 July 2021; pp. 8748–8763.

24.　Guo, H.; Cao, J.; Zhang, Y.; Guo, J.; Li, J. Rumor detection with hierarchical social attention network. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 943–951.

25.　Chen, Y. Convolutional Neural Network for Sentence Classification. Master's Thesis, University of Waterloo, Waterloo, ON, Canada, 2015.

26.　Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; Chi, E.H. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 1930–1939.

27.　Qin, Z.; Cheng, Y.; Zhao, Z.; Chen, Z.; Metzler, D.; Qin, J. Multitask mixture of sequential experts for user activity streams. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual, 6–10 July 2020; pp. 3083–3091.

28.　Nan, Q.; Cao, J.; Zhu, Y.; Wang, Y.; Li, J. MDFEND: Multi-domain fake news detection. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Gold Coast, Australia, 1–5 November 2021; pp. 3343–3347.

29.　Wu, K.; Yang, S.; Zhu, K.Q. False rumors detection on sina weibo by propagation structures. In Proceedings of the 2015 IEEE 31st International Conference on Data Engineering, Seoul, Republic of Korea, 13–17 April 2015; pp. 651–662.

30.　Ajao, O.; Bhowmik, D.; Zargari, S. Sentiment aware fake news detection on online social networks. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2507–2511.

31.　Jiang, W.; Chen, B.; He, L.; Bai, Y.; Qiu, X. Features of rumor spreading on wechat moments. In Proceedings of the Web Technologies and Applications: APWeb 2016 Workshops, WDMA, GAP, and SDMA, Suzhou, China, 23–25 September 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 217–227.

32.　Ma, J.; Gao, W.; Wei, Z.; Lu, Y.; Wong, K.F. Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 18–23 October 2015; pp. 1751–1754.

33. Dai, E.; Sun, Y.; Wang, S. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In Proceedings of the International AAAI Conference on Web and Social Media, Atlanta, GA, USA, 8 June 2020; Volume 14, pp. 853–862.

34. Yu, F.; Liu, Q.; Wu, S.; Wang, L.; Tan, T. A Convolutional Approach for Misinformation Identification. In Proceedings of the IJCAI, Melbourne, Australia, 19–25 August 2017; pp. 3901–3907.

35. Chen, T.; Xu, R.; He, Y.; Wang, X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Syst. Appl.* **2017**, *72*, 221–230. [CrossRef]

36. Singh, J.P.; Kumar, A.; Rana, N.P.; Dwivedi, Y.K. Attention-based LSTM network for rumor veracity estimation of tweets. *Inf. Syst. Front.* **2022**, *24*, 459–474. [CrossRef]

37. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.

38. Qazvinian, V.; Rosengren, E.; Radev, D.; Mei, Q. Rumor has it: Identifying misinformation in microblogs. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011; pp. 1589–1599.

39. Morris, M.R.; Counts, S.; Roseway, A.; Hoff, A.; Schwarz, J. Tweeting is believing? Understanding microblog credibility perceptions. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, Seattle, WA, USA, 11–15 February 2012; pp. 441–450.

40. Suzuki, Y. A credibility assessment for message streams on microblogs. In Proceedings of the 2010 International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, Fukuoka, Japan, 4–6 November 2010; pp. 527–530.

41. Mohammad, S.M.; Sobhani, P.; Kiritchenko, S. Stance and sentiment in tweets. *ACM Trans. Internet Technol. (TOIT)* **2017**, *17*, 1–23. [CrossRef]

42. Liang, G.; He, W.; Xu, C.; Chen, L.; Zeng, J. Rumor identification in microblogging systems based on users' behavior. *IEEE Trans. Comput. Soc. Syst.* **2015**, *2*, 99–108. [CrossRef]

43. Yang, S.; Shu, K.; Wang, S.; Gu, R.; Wu, F.; Liu, H. Unsupervised fake news detection on social media: A generative approach. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 5644–5651.

44. Gao, M.; Chen, F. Credibility evaluating method of Chinese microblog based on information fusion. *J. Comput. Appl.* **2016**, *36*, 2071.

45. Bazmi, P.; Asadpour, M.; Shakery, A. Multi-view co-attention network for fake news detection by modeling topic-specific user and news source credibility. *Inf. Process. Manag.* **2023**, *60*, 103146. [CrossRef]

46. Hu, L.; Chen, Z.; Yin, Z.Z.J.; Nie, L. Causal Inference for Leveraging Image-text Matching Bias in Multi-modal Fake News Detection. *IEEE Trans. Knowl. Data Eng.* **2022**. [CrossRef]

47. Xiong, S.; Zhang, G.; Batra, V.; Xi, L.; Shi, L.; Liu, L. TRIMOON: Two-Round Inconsistency-based Multi-modal fusion Network for fake news detection. *Inf. Fusion* **2023**, *93*, 150–158. .. [CrossRef]

48. Li, Q.; Zhang, Q.; Si, L. 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, Florence, Italy, 28 July–2 August 2019; pp. 1173–1179.

49. Ma, J.; Gao, W.; Wong, K.F. Detect rumor and stance jointly by neural multi-task learning. In Proceedings of the Companion Proceedings of the the Web Conference 2018, Lyon, France, 23–27 April 2018; pp. 585–593.

50. Zhu, Y.; Sheng, Q.; Cao, J.; Li, S.; Wang, D.; Zhuang, F. Generalizing to the Future: Mitigating Entity Bias in Fake News Detection. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), New York, NY, USA, 11–15 July 2022; pp. 2120–2125. [CrossRef]

51. Zhao, J.; Du, B.; Sun, L.; Zhuang, F.; Lv, W.; Xiong, H. Multiple relational attention network for multi-task learning. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 1123–1131.

52. Zhu, Y.; Liu, Y.; Xie, R.; Zhuang, F.; Hao, X.; Ge, K.; Zhang, X.; Lin, L.; Cao, J. Learning to expand audience via meta hybrid experts and critics for recommendation and advertising. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; pp. 4005–4013.

53. Zhu, Y.; Zhuang, F.; Wang, D. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 5989–5996.

54. Sun, M. weibo_senti_100k and THUCNews 2022. Available online: https://ieee-dataport.org/documents/weibosenti100k-and-thucnews (accessed on 2 August 2022).

55. Sina Weibo. Available online: http://www.weibo.com (accessed on 15 April 2022).

56. Newsverify. Available online: https://www.newsverify.com/ (accessed on 23 April 2022).

57. WeiboService. Available online: http://service.account.weibo.com/ (accessed on 15 April 2022).

58. Vosoughi, S.; Roy, D.K.; Aral, S. The spread of true and false news online. *Science* **2018**, *359*, 1146–1151. [CrossRef]

59. Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In Proceedings of the Machine Learning: ECML-98: 10th European Conference on Machine Learning, Chemnitz, Germany, 21–23 April 1998; Springer: Berlin/Heidelberg, Germany, 2005; pp. 137–142.

60. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
61. Mikolov, T.; Chen, K.; Corrado, G.S.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the International Conference on Learning Representations, Scottsdale, AZ, USA, 2–4 May 2013.
62. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning PMLR, Beijing, China, 21–26 June 2014; pp. 1188–1196.
63. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
64. Johnson, G. Google Translate http://translate.google.com . *Tech. Serv. Q.* **2012**, *29*, 165. [CrossRef]
65. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* **2020**, *21*, 5485–5551.