

DEAR: Disentangled Event-Agnostic Representation Learning for Early Fake News Detection

Xiao Pu, Hao Wu, Xiuli Bi*, Yu Wu, Xinbo Gao

Chongqing University of Posts and Telecommunications, China

puxiao@cqupt.edu.cn, wuhao_stu_cqupt@outlook.com,

bixl@cqupt.edu.cn, wuyu@cqupt.edu.cn, gaoxb@cqupt.edu.cn

Abstract

Detecting fake news early is challenging due to the absence of labeled articles for emerging events in training data. To address this, we propose a **Disentangled Event-Agnostic Representation (DEAR)** learning approach. Our method begins with a BERT-based adaptive multi-grained semantic encoder that captures hierarchical and comprehensive textual representations of the input news content. To effectively separate latent authenticity-related and event-specific knowledge within the news content, we employ a disentanglement architecture. To further enhance the decoupling effect, we introduce a cross-perturbation mechanism that perturbs authenticity-related representation with the event-specific one, and vice versa, deriving a robust and discerning authenticity-related signal. Additionally, we implement a refinement learning scheme to minimize potential interactions between two decoupled representations, ensuring that the authenticity signal remains strong and unaffected by event-specific details. Experimental results demonstrate that our approach effectively mitigates the impact of event-specific influence, outperforming state-of-the-art methods. In particular, it achieves a 6.0% improvement in accuracy on the PHEME dataset over MDDA, a similar approach that decouples latent content and style knowledge, in scenarios involving articles from unseen events different from the topics of the training set.

1 Introduction

Social media platforms provide a convenient way for the public to create, distribute, and absorb a wide range of information. However, the authenticity of information shared on these platforms has become a growing concern. When users cannot

verify the authenticity of the content they share, there is a significant risk of spreading fake information, leading to the widespread dissemination across platforms and posing serious consequences.

Advances in natural language processing (NLP) have facilitated the development of fake news detection methods (as described in Section 4.1), which have shown significant performance in controlled scenarios, as illustrated in Figure 1 (b). However, on social media, sudden explosive events often lead to a flood of related news. In the early stage of news dissemination, resources related to newly emerging events are usually limited, and authenticity annotations for those news articles are often unavailable. This makes it challenging to quickly train new fake news detection models. Meanwhile, detection models trained on existing annotated data typically perform poorly on emerging “unseen” events during inference, as shown in Figure 1(c). Therefore, it highlights the challenge of quickly detecting and blocking fake news articles about these newly emerging events before they spread widely.

Recent methods of fake news detection address this challenge by treating distinct events as various domains and leveraging domain adaptation techniques (as described in Section 4.2). These methods aim to bridge the semantic gaps between events by enhancing their correlations. However, they require a portion of data from the target domain during training, making them inapplicable for the early fake news detection scenario with “unseen” events. Moreover, since newly emerging events on social media appear frequently, it is impractical to adapt the model each time a new event occurs. Therefore, a model is needed to identify fake news immediately from most newly emerging “unseen” events.

Based on this analysis, we assume that a news article contains two key latent attributes: authenticity-related knowledge, which determines

*Corresponding author.

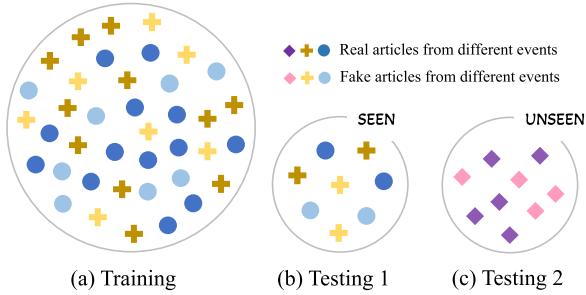


Figure 1: Illustration of different fake news detection scenarios: (a) The training set, composed of news articles from two existing events. (b) This represents a type of testing set including articles related to the same events as the training set, referred to as “seen” events. In contrast, (c) illustrates another type of testing set containing articles about new events not present in training, showcasing a detection scenario for “unseen” events.

if it is fake or real, and event-specific knowledge, which varies significantly among different events. We explore the performance of fake news detection for newly “unseen” events by utilizing the latent authenticity-related knowledge and mitigating the impact of the event-specific knowledge. Building on these insights from our exploration, we propose a **Disentangled Event-Agnostic Representation** (DEAR) learning approach to address the challenges of early fake news detection for newly “unseen” events (explained in Section 2). It begins with an adaptive multi-grained semantic encoder for contextual representation extraction. Next, a disentanglement architecture is utilized to decouple the latent authenticity-related and event-specific representations of news content. To enhance the disentanglement effect, we introduce a cross-perturbation mechanism that perturbs the generated authenticity-related representation with the event-specific one, and vice versa. This directs the discriminators to focus on each corresponding knowledge, thereby improving their robustness. Finally, a refinement learning scheme is employed to further reduce potential interactions between the two decoupled representations, ensuring the authenticity-related signal remains strong and unaffected by event-specific details. We demonstrate the effectiveness of our approach through comprehensive experiments conducted on datasets, showcasing its superiority over established methods (outlined in Section 3).

2 The DEAR Methodology

2.1 Preliminary Analysis

Existing fake news detection methods struggle with identifying fake news related to events not encountered during training. To analyze this challenge, we conduct a preliminary detection experiment using a multi-event configuration. In this setup, we train the model on news articles related to two distinct events, namely, the U.S. Election and COVID-19, and test it on articles about the “Ferguson Unrest” event. For the U.S. Election, we extract approximately 600 articles containing keywords such as “election” and “president” from the PolitiFact dataset (Shu et al., 2018). For COVID-19, we randomly select around 600 articles with balanced authenticity labels from the COVID dataset (Du et al., 2021), which contains articles related to COVID-19. For the Ferguson Unrest event, we randomly select around 600 articles related to this event from the PHEME dataset (Kochkina et al., 2018), which contains articles from multiple events with both event and authenticity labels.

In this experiment, we fine-tune BERT (Devlin et al., 2019) using articles from the U.S. Election and COVID-19 events for training. We then visualize the semantic representations learned by BERT for news articles from both training and testing events with t-SNE, as shown in Figure 2(a). The representations of various events exhibit notable disparities, indicating significant semantic differences between distinct events. Consequently, the authenticity detection performance on the unseen Ferguson Unrest event is notably compromised.

To uncover the latent factors influencing such distribution disparities, we introduce two separate branches over BERT’s embeddings, aiming to extract the authenticity-related and event-specific knowledge separately. Each branch comprises a single-layer MLP network as a generator, denoted as G , and a two-layer MLP network as a discriminator, denoted as D . To decouple event-specific and authenticity-related information from the BERT semantic embeddings, the generator G_e in the event branch is guided by the corresponding discriminator D_e using event labels (“COVID-19” or “U.S. Election”), while the generator G_c along with the discriminator D_c in the authenticity branch is trained according to the fake/real labels. The representations obtained

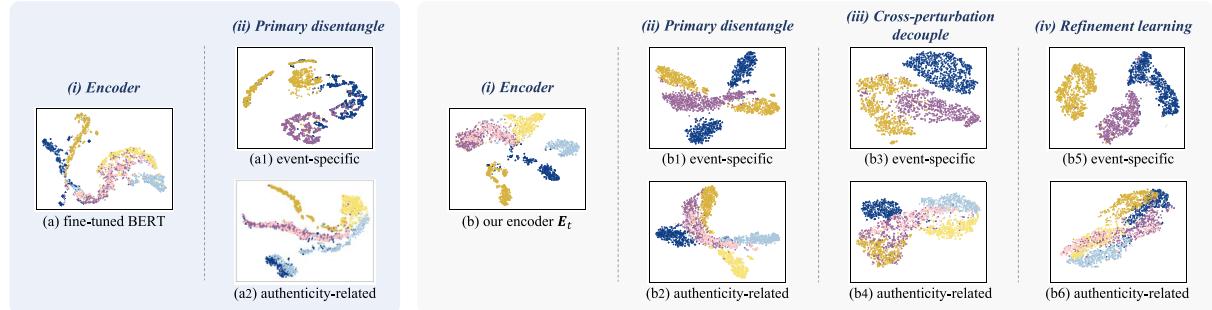


Figure 2: t-SNE visualization on distributions generated by various models: (a) and (b) show the embeddings of news articles learned from different encoders, while (ii)-(iv) illustrate the representations of news articles learned from event and authenticity generators across three distinct models. For the training set, we select news articles from two distinct events: U.S. Election event (yellow) and COVID-19 event (blue). The testing set comprises news from Ferguson Unrest event (purple). In the visualization, darker shades indicate real news articles while lighter shades represent fake ones.

from the event and authenticity generators are visualized via t-SNE in Figure 2(a1) and (a2).

Compared with the original embeddings from BERT in Figure 2(a), the representations learned from the event generator in Figure 2(a1) cluster the events to some extent and slightly reduce the fake/real distance between events. Conversely, the representations learned from the authenticity generator shown in Figure 2(a2) marginally reduce event gaps and slightly increase the fake/real distances.

Based on these observations, we hypothesize that latent authenticity-related knowledge, independent of specific events, can reduce the semantic distance among events and enhance the performance of detecting fake news in unseen events. To improve the detection performance on unseen events, we prioritize learning authenticity-related knowledge while mitigating the impact of event-specific knowledge.

2.2 Proposed Method

The proposed method is outlined in Figure 3. It begins with an adaptive multi-grained semantic encoder to extract comprehensive representations of each news article. Then, a cross-perturbation decoupling mechanism is employed to extract authenticity-related and event-specific knowledge. The authenticity-related knowledge serves as an indicator of news authenticity, facilitating more effective detection of fake news across various events and scenarios, including those from unseen events. Additionally, a refinement step is included to further enhance the decoupled

authenticity-related knowledge by filtering out any residual event-specific information.

2.2.1 Adaptive Multi-Grained Semantic Encoder

In the field of NLP, one of the fundamental and crucial challenges is to learn a comprehensive representation of a given text, and fake news detection is no exception. Conventional representation learning mechanisms often involve either fine-grained word-level learning or abstracted document-level learning. Local-based word-level representations, which express the embedding of each word in the document as semantic matrices, may capture detailed information but lack a broad view, potentially leading the model to focus on specific keywords. In contrast, global-based document-level representations summarize the semantic knowledge of the entire document as a single vector, providing a summarized understanding but potentially overlook important details.

To capture latent authenticity-related knowledge from given real and fake articles, we propose an adaptive multi-grained semantic encoder E_t , as shown in Figure 4. This encoder aggregates both fine-grained and coarse information, encompassing both global (document-level) and local (word-level) representations, as well as their interactions within a given news article. In E_t , we first extract the semantic embedding of the [CLS] token from BERT as the global-based representation $T_g \in \mathbb{R}^{1 \times D}$. We also obtain the overall semantic

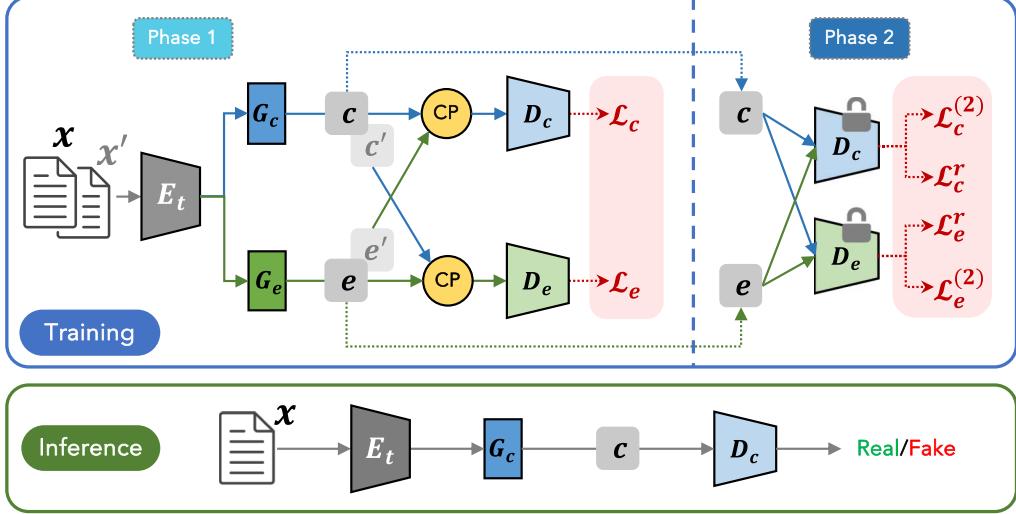


Figure 3: The overview of DEAR methodology, composed of three principal modules: the adaptive multi-grained semantic encoder E_t , the authenticity/event generators $G_{c/e}$ (each of them constructed with a single-layer MLP network), and the authenticity/event discriminator $D_{c/e}$ (each of them designed with a two-layer MLP network). Dual-phase training is utilized to learn the authenticity-related representation that serves as a reliable and robust signal for fake news detection.

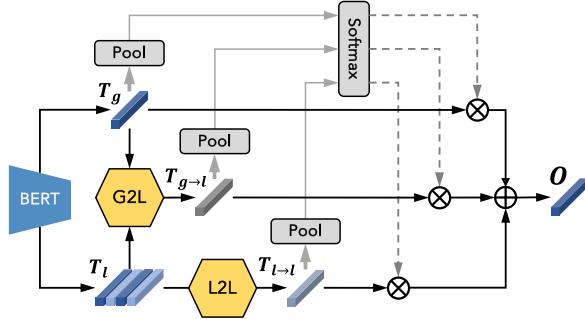


Figure 4: The architecture of the proposed adaptive multi-grained semantic encoder.

embedding from BERT as the local-based representation $\mathbf{T}_l \in \mathbb{R}^{L \times D}$. These embeddings serve as the initial representations of the input text.

To capture the interactions among different levels of granularity in the encoded semantic information, we introduce a coherence measurement between global-to-local representations:

$$\mathbf{T}_{g \rightarrow l} = \sum_{i=1}^L \left(\left\| \sum_{h=1}^H \sigma \left(\frac{[\mathbf{T}_g][\mathbf{T}_l]^\top}{\sqrt{d_r}} \right) \right\| \right) \times \mathbf{T}_l^{(i)}, \quad (1)$$

and local-to-local representations:

$$\mathbf{T}_{l \rightarrow l} = \sum_{i=1}^L \left(\left\| \sum_{h=1}^H \sigma \left(\frac{[\mathbf{T}_l][\mathbf{T}_l]^\top}{\sqrt{d_r}} \right) \right\| \right) \times \mathbf{T}_l^{(i)}, \quad (2)$$

where L is the maximal length of the given news article, H is the number of heads of the multi-head attention, and σ denotes the Softmax operation.

To construct the final text representation \mathbf{o} from various perspectives, we employ an attentive aggregation mechanism. We first compute the average pooling of three distinct vectors \mathbf{T}_g , $\mathbf{T}_{g \rightarrow l}$ and $\mathbf{T}_{l \rightarrow l}$. Then, we apply Softmax to determine the weights $\{a_g, a_{g \rightarrow l}, a_{l \rightarrow l}\}$. Finally, we obtain the final representation by combining the obtained multi-view representations for enhanced text representation:

$$\mathbf{o} = a_g \mathbf{T}_g + a_{g \rightarrow l} \mathbf{T}_{g \rightarrow l} + a_{l \rightarrow l} \mathbf{T}_{l \rightarrow l}. \quad (3)$$

Based on the preliminary detection configuration described in Section 2.1, we replace the BERT encoder shown in Figure 2(a) with our encoder and visualize the distribution of our encoder's output representations illustrated in Figure 2(b). It can be seen that the proposed encoder E_t enlarges the overall distance between fake and real news articles by capturing the hierarchical differences between fake and real samples. Additionally, we embed E_t into the primary disentanglement framework and extract the decoupled event-specific (b1) and authenticity-related knowledge (b2) for comparison. The comparison shows that the generated

event-specific representations based on our encoder tighten the news distribution for each event, resulting in more apparent clustering (b1 vs. a1 in Figure 2). For authenticity-related knowledge comparison (b2 vs. a2 in Figure 2), our encoder’s generated features show more overlapping across various events.

2.2.2 Cross-Perturbation Decoupling

In Section 2.1, we mention two branches to generate and discriminate authenticity-related and event-specific representations over the semantic embeddings from BERT. Our analysis of the experimental results illustrated in Figure 2(a)–(a2) led us to conclude that latent authenticity-related knowledge, yet independent of specific events, shows promise for detecting fake news in unseen events. Based on this conclusion, we propose a novel cross-perturbation (CP) mechanism that randomly mixes authenticity-related and event-specific knowledge from two distinct news articles. The CP mechanism can enhance the learning process of authenticity-related features by introducing event-specific perturbations, and vice versa.

Given an input news article x , we extract \mathbf{o} by the proposed encoder \mathbf{E}_t . This representation is then fed into the authenticity generator \mathbf{G}_c and event generator \mathbf{G}_e , each constructed with a single-layer MLP network. From these generators, we obtain \mathbf{c} and \mathbf{e} , representing the authenticity-related and event-specific aspects for \mathbf{o} , respectively. To introduce perturbations, we also extract authenticity-related and event-specific representations (\mathbf{c}' and \mathbf{e}') from a randomly sampled news article x' .

For the authenticity branch, we merge the authenticity-related representation \mathbf{c} with the event-specific knowledge by:

$$\tilde{\mathbf{c}} = \alpha \times \mathbf{c} + (1 - \alpha) \times \text{AdaIN}(\mathbf{e}', \mathbf{c}), \quad (4)$$

where $\text{AdaIN}(., .)$ is the Adaptive Instance Normalization (Huang and Belongie, 2017), a feature fusion technique from image style transfer, α is an interpolation weight sampled uniformly from $[0, 1]$, and $\tilde{\mathbf{c}}$ denotes the authenticity-related representation with injected event-specific noise. $\tilde{\mathbf{c}}$ is then fed into the authenticity discriminator \mathbf{D}_c , constructed with a two-layer MLP network. Additionally, $\tilde{\mathbf{e}}$, which encapsulates authenticity-related information from the selected

sample x' , is also fed into the authenticity discriminator to assist the fake news detection based on the authenticity label of x' . This process is trained with the loss \mathcal{L}_c for binary fake news detection:

$$\mathcal{L}_c = - \left[\mathbb{E}_{(\tilde{\mathbf{c}}, y)} \sum_{i=1}^K y_i \log \mathbf{D}_c(\tilde{\mathbf{c}}_i) + \mathbb{E}_{(\tilde{\mathbf{e}}, y')} \sum_{i=1}^K y'_i \log \mathbf{D}_c(\tilde{\mathbf{e}}_i) \right], \quad (5)$$

where K denotes the number of class labels, y' is the authenticity label of the selected sample x' .

For the event branch, we perturb the event-specific representations \mathbf{e} with authenticity-related knowledge according to:

$$\tilde{\mathbf{e}} = \beta \times \mathbf{e} + (1 - \beta) \times \text{AdaIN}(\mathbf{c}', \mathbf{e}), \quad (6)$$

where β is an interpolation weight uniformly sampled from $[0, 1]$, and $\tilde{\mathbf{e}}$ represents the event-specific representation perturbed by the authenticity-related noise. The obtained $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{c}}$ are then fed into the event discriminator \mathbf{D}_e , to predict the event labels for the samples x and x' , respectively.

The objective of the network is to predict the specific event label by minimizing the event-centric loss \mathcal{L}_e defined as:

$$\mathcal{L}_e = - \left[\mathbb{E}_{(\tilde{\mathbf{e}}, d)} \sum_{i=1}^D d_i \log \mathbf{D}_e(\tilde{\mathbf{e}}_i) + \mathbb{E}_{(\tilde{\mathbf{c}}, d')} \sum_{i=1}^D d'_i \log \mathbf{D}_e(\tilde{\mathbf{c}}_i) \right], \quad (7)$$

where D denotes the total number of events referenced by the articles in the training set.

By perturbing authenticity-related representations with event-specific information during training, the authenticity discriminator is compelled to prioritize authenticity-related knowledge over the event-specific details for more accurate fake news detection. Furthermore, this CP mechanism enhances the diversity of input samples, enriching the training dataset. The overall loss objective of the first training phase is defined as:

$$\mathcal{L}^{(1)} = \mathcal{L}_c + \lambda_1 \cdot \mathcal{L}_e, \quad (8)$$

where λ_1 is a trade-off hyper-parameter.

To understand the effectiveness of our CP mechanism within the disentanglement framework, we visualize the t-SNE distribution of generated event-specific and authenticity-related representations learned from the corresponding generators during the inference process in Figure 2 (b3) and (b4).

Our model, which includes the CP mechanism, exhibits two clearly decoupled representations, as shown in (b3) and (b4) of Figure 2, compared to the primary disentanglement framework shown in Figure 2(b1) and (b2). The authenticity-related representation captures the latent invariant knowledge across various events that determine authenticity. The authenticity-related distribution in Figure 2(b4) effectively bridges the gap between articles from distinct events in both fake and real clusters while maintaining the distance between the fake and real clusters. On the other hand, the event-specific representation clearly categorizes the three distinct events, as shown in Figure 2(b3). In contrast, the primary disentanglement framework shows significant overlap between the two decoupled representations, with the event-specific representation (b1) inadequately separating true and false samples. This indicates a failure to properly decouple the authenticity-related and event-specific knowledge.

These comparisons highlight how the proposed CP mechanism effectively minimizes the overlap between the two decoupled representations, enhancing the distinction of each representation.

2.2.3 Refinement Learning

To further enhance the disentanglement capability, we introduce a refinement learning strategy in the second phase of training, inspired by adversarial learning (Goodfellow et al., 2014). This strategy aims to ensure that the decoupled authenticity-related representation does not include any event-specific knowledge identifiable by a proficiently trained event discriminator, and vice versa. In this phase, we freeze both the authenticity and event discriminators while refining the two generators.

We consider two refinement scenarios simultaneously: In the first scenario, the event-specific and authenticity-related representations are fed into their corresponding discriminators, aiming for the detection and classification results to be

as accurate as possible. The loss function for this scenario is defined as:

$$\begin{aligned} \mathcal{L}_{y,d}^{(2)} &= \mathcal{L}_c^{(2)} + \mathcal{L}_e^{(2)} \\ &= - \left[\mathbb{E}_{(c,y)} \sum_{i=1}^K y_i \log \mathbf{D}_c(\mathbf{c}_i) + \right. \\ &\quad \left. \mathbb{E}_{(e,d)} \sum_{i=1}^D d_i \log \mathbf{D}_e(\mathbf{e}_i) \right]. \quad (9) \end{aligned}$$

In the second scenario, two types of representations are fed into the opposing discriminators, where the detection and classification results are expected to be randomized. This process is optimized according to:

$$\begin{aligned} \mathcal{L}_{y^r,d^r}^{(2)} &= \mathcal{L}_c^r + \mathcal{L}_e^r \\ &= - \left[\mathbb{E}_{(e,y^r)} \sum_{i=1}^K y_i^r \log \mathbf{D}_c(\mathbf{e}_i) + \right. \\ &\quad \left. \mathbb{E}_{(c,d^r)} \sum_{i=1}^D d_i^r \log \mathbf{D}_e(\mathbf{c}_i) \right], \quad (10) \end{aligned}$$

where y^r and d^r represent random authenticity and event labels sampled from a uniform distribution. The overall loss objective for the second training stage is therefore defined as:

$$\mathcal{L}^{(2)} = (1 - \lambda_2) \cdot \mathcal{L}_{y^r,d^r}^{(2)} + \lambda_2 \cdot \mathcal{L}_{y,d}^{(2)}, \quad (11)$$

where λ_2 is the hyper-parameter governing the trade-off between the loss items. The second training phase aims to refine the disentangled representations by filtering out the potential overlapping knowledge between authenticity-related and event-specific representations.

We present the t-SNE distributions of the authenticity-related and event-specific representations learned from the CP decoupling model with refinement learning, as illustrated in Figure 2 (b6) and (b5). Compared with the distributions in (iii), which lack the second phase strategy, the event-specific features from the refinement learning process (b5) are tightly clustered around independent event clusters. Meanwhile, the authenticity-related representations, as shown in Figure 2(b6), exhibit more overlaps compared to the model without refinement learning, shown in Figure 2(b4). This comparison demonstrates that the refinement learning strategy enhances the independence of each decoupled representation and reduces the potential inter-correlation between the two.

Scenario	Datasets	Events	Description	Topic	All	Fake	Real
In-Topic Detection (ITD)	PHEME (Kochkina et al., 2018)	C_h	Charlie Hebdo: Terrorist attack on the French satirical magazine in Paris, resulting in 12 deaths.	Social Unrest	2,079	458	1,621
		S_y	Sydney Siege: It was a 16-hour hostage crisis at the Lindt Café in Sydney, resulting in three deaths, including the gunman.		1,221	522	699
		F_e	Ferguson Unrest: It involved protests and riots in Ferguson, Missouri, after the fatal shooting of Michael Brown, an unarmed Black teenager, by a police officer on 2014.		1,143	284	859
		O_t	Ottawa Shooting: The shooting involved a gunman killing a soldier at the National War Memorial before being shot dead after storming the Canadian Parliament.		890	470	420
Cross-Topic Detection (CTD)	PolitiFact (Shu et al., 2018)	P_o	Comprised of multiple events related to politics, such as U.S. Election and policy debates.	Politics	948	420	528
	GossipCop (Shu et al., 2018)	G_o	Comprised of multiple events related to the gossip topic, such as Celebrity Death Hoaxes and entertainment stories	Gossip	9,947	4,947	5,000
	COVID (Du et al., 2021)	C_o	Comprised of tweets related to COVID-19.	Health	6,067	1,317	4,750

Table 1: Statistics of selected public datasets.

Upon completing the dual-phase training, only the pair of authenticity generator and discriminator is employed during the testing stage of fake news detection. This setup outputs the binary detection label based on the disentangled authenticity-related representation, effectively mitigating event-specific information. The CP mechanism and Refinement learning strategy are involved solely during the training process and are excluded during inference.

3 Experiments

3.1 Experimental Settings

Datasets We use four datasets as listed in Table 1, to simulate two types of unseen-event detection scenarios:

- **In-Topic Detection (ITD):** In this scenario, we utilize the PHEME dataset, which includes thousands of claims related to four different events centered around a similar topic of social unrest. We use three of the events as the resource for training and the remaining event as the target for testing, resulting in four different combinations.

- **Cross-Topic Detection (CTD):** To further challenge our proposed approach, we incorporate news articles from three other datasets, i.e., PolitiFact, GossipCop, and COVID, which cover distinctly different events. In

this scenario, we use two of these datasets as the resources for training and evaluate the trained model on the remaining dataset, resulting in three different combinations.

Selection of Comparison Methods We compare our method with two types of baselines, namely, content-centric and domain-adaptive approaches. The content-centric baselines include TextCNN (Kim, 2014) and RoBERTa (Liu et al., 2019), both of which leverage content knowledge for text classification. TextCNN is a convolutional network robust in various text classification tasks, while RoBERTa, a variant of the pre-trained transformer, generates embeddings of the [CLS] token for detection. The domain-adaptive baselines include EANN (Wang et al., 2018), MDDA (Zhang et al., 2021), Fish (Shi et al., 2022), and metaAdapt (Yue et al., 2023). EANN learns event-agnostic features using a TextCNN for text representation and an event discriminator for adversarial learning. MDDA disentangles the representation into content- and style-based branches, utilizing only style knowledge for detection. As a multi-modal framework, we consider its textual branch for comparison. Fish introduces a gradient-based framework for domain generalization, featuring an adaptive mechanism for handling different domains. MetaAdapt employs meta-training to explore optimal parameters, enabling rapid adaptation to unseen tasks without examples. To ensure

From	Cross-Topic Detection (CTD)		
	\mathcal{G}_o	\mathcal{P}_o	\mathcal{C}_o
BERT	0.0192	0.0164	0.0126
E_t	0.7256	1.0128	0.9069
From	In-Topic Detection (ITD)		
	\mathcal{C}_h	\mathcal{O}_t	\mathcal{S}_y
BERT	0.0271	0.0198	0.0173
E_t	0.6184	0.7183	0.7294

Table 2: Wasserstein distance between the embeddings of real vs. fake news is examined in both *CTD* and *ITD* scenarios, comparing between the original BERT and our proposed encoder E_t . The E_t increases the separation between distributions by a large margin on all datasets.

a fair comparison, we train these frameworks using the same datasets as ours.

Implementation Details In our experiment, we use the Adam optimizer with a learning rate of 2e-5 and a weight decay of 0.01. The training batch size is set to 32. The entire training process takes around 40 epochs to converge on both *ITD* and *CTD* scenarios. All experiments are conducted on a single NVIDIA GeForce RTX 3090 GPU.¹

Evaluation Metrics We select *Acc* and F_1 as the major evaluation metrics to measure the performance of various approaches, which are commonly utilized in the context of fake news detection. Furthermore, Wasserstein distance is also used as a metric to measure the distance between two distributions, such as real vs. fake news. We also perform t-SNE visualization as a visual evaluation mechanism in our experiments.

3.2 Discussion of the Proposed Components

Before presenting the final detection performance, we first validate the effectiveness of each proposed component.

Effectiveness of E_t We evaluate the encoder’s ability to differentiate between fake and real news in both *CTD* and *ITD* scenarios. Specifically, for both BERT and our encoder E_t , we calculate the Wasserstein distances between the centers of fake and real representations for both detection scenarios, as shown in Table 2. It is clear

¹Our source code will be publicly available at <https://github.com/PuXiao06/DEAR>.

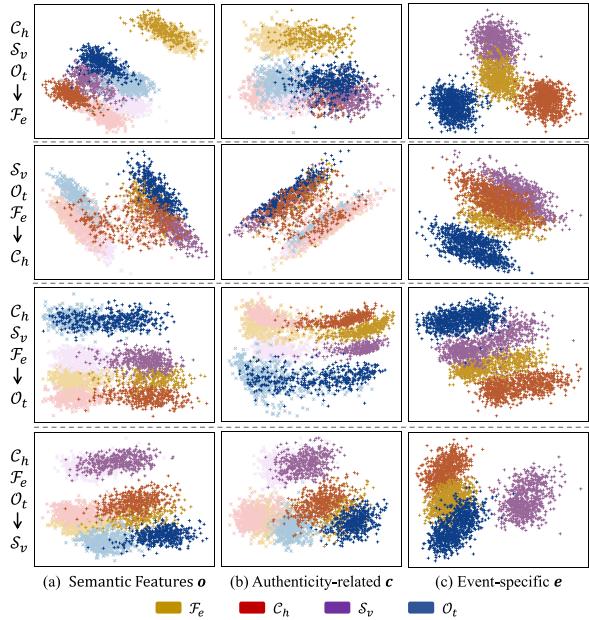


Figure 5: T-SNE visualization of the representation distributions via different event combinations evaluated on PHEME corpus, trained on a subset of three events and tasked with detecting fake news on the remaining one, as an unseen event. Different colors represent different events, while dark and light shades distinguish between real and fake samples.

that the proposed E_t effectively enlarges the distance between fake and real news, indicating its effectiveness in understanding fine-grained and comprehensive distinctions and correlations among news instances. This comparison substantiates the network’s capability to discern and leverage the contextual information, contributing to improved detection capability.

Effectiveness of CP Decoupling Mechanism

To gain insights about the effectiveness of the proposed CP decoupling mechanism, we inspect the development of representation distributions before and after feature disentanglement. We compare the t-SNE distributions of the original representation o , the decoupled authenticity-related c , and the decoupled event-specific e in *ITD* scenario using the PHEME dataset, as illustrated in Figure 5.

As shown in the first column of Figure 5, the original representations o corresponding to real and fake samples in training set are well separated. However, this separation is not so clear for the testing data, which features an unseen event (yellow color). With the introduction of CP decoupling, represented in the second column (b) of Figure 5, the gap between real and fake articles in the decoupled authenticity-related representations c is

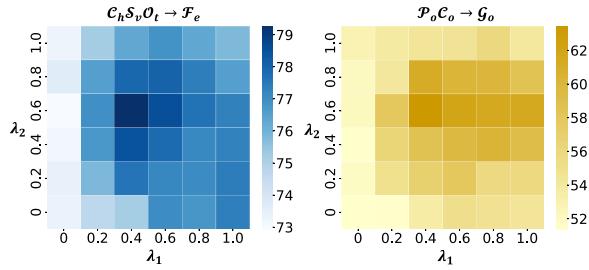


Figure 6: Comparison of detection accuracy results on multiple datasets with different λ_1 and λ_2 .

further enlarged, while the gap between different events is significantly reduced. Simultaneously, the decoupled event-specific representations e show clear isolation between distinct events, precisely capturing the unique knowledge characteristics associated with each event. This analysis confirms that the proposed CP decoupling mechanism effectively mitigates the event-specific knowledge from the original semantic representation, enhancing the model’s capability to discern real from fake samples by utilizing the authenticity-related information exclusively.

Hyper-Parameter Selection To confirm the optimal hyper-parameter values, we conduct experiments using one dataset combination from each of the *ITD* and *CTD* scenarios. Among the parameters, the scaling factors λ_1 and λ_2 play crucial roles in determining the relative weight of the event branch and refinement learning component, respectively. Specifically, λ_1 influences the trade-off between authenticity detection and event classification for the objective $\mathcal{L}^{(1)}$, while λ_2 impacts the trade-off between the two loss items in the loss function $\mathcal{L}^{(2)}$. We test different combinations of hyper-parameters, with values from $\{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ in each iteration. The detection accuracy of $\mathcal{C}_h \mathcal{S}_v \mathcal{O}_t \rightarrow \mathcal{F}_e$ from *ITD* scenario and $\mathcal{P}_o \mathcal{C}_o \rightarrow \mathcal{G}_o$ from *CTD* scenario are shown in Figure 6. For both cases, the combination of $\lambda_1 = 0.4$ and $\lambda_2 = 0.6$ results in the optimal detection performance. Hence, these two values are used throughout our experiment.

3.3 Quantitative Detection Evaluation

Evaluation of the *ITD* Scenario To evaluate the proposed approach against comparative methods, we first present in Table 3 the overall accuracy for the *ITD* scenario, where the target events are different from the source events but share a similar topic. It is clear that the proposed

approach outperforms all competitive approaches with significant improvements. Specifically, we achieve relative gains of 6.0% and 5.5% in accuracy and F_1 , respectively, for the combination of $\mathcal{C}_h \mathcal{F}_e \mathcal{O}_t \rightarrow \mathcal{S}_y$ over the recent MDDA, a typical method utilizing a similar disentanglement mechanism for managing style and content knowledge discrepancies.

Evaluation of the *CTD* Scenario To further challenge our approach, we evaluate its performance in the *CTD* scenario, where the topics of targeted unseen events are distinct from those of the source events, with results shown in Table 4. It can be seen that the proposed approach consistently outperforms the competitive approaches with a clear margin. Especially for the combination of $\mathcal{P}_o \mathcal{C}_o \rightarrow \mathcal{G}_o$, we achieve a significant accuracy gain over 7.71% compared to the MetaAdapt (Yue et al., 2023), which uses a meta-training strategy for rapid adaptation to target data.

3.4 Ablation Study

To validate the contribution of each component of the proposed approach, we conduct a set of ablation experiments on both detection scenarios, employing different configuration combinations within our model. As shown in Figure 7, the proposed encoder E_t consistently demonstrates its advantages over BERT in extracting more hierarchical knowledge, enabling enhanced comprehension of the authenticity-related information. The detection performance is further improved when the CP decoupling mechanism is included (labeled as “+CP decouple”), which effectively mitigates the influence of the event-specific noise by preserving the pure authenticity-related knowledge for real/fake detection. Utilizing the refinement learning based on the CP decoupling mechanism (labeled as “+refine”) significantly increases the detection accuracy by enhancing the robustness of both discriminators. This component plays a crucial role in enhancing the disentanglement task by further filtering out possible interactive knowledge between the authenticity-related and event-specific representations. This analysis confirms the contribution of each proposed key component in our approach, with the cross-perturbation decoupling mechanism yielding the highest improvement in the detection performance.

Method	$\mathcal{C}_h \mathcal{S}_y \mathcal{O}_t \rightarrow \mathcal{F}_e$		$\mathcal{C}_h \mathcal{F}_e \mathcal{O}_t \rightarrow \mathcal{S}_y$		$\mathcal{S}_y \mathcal{F}_e \mathcal{O}_t \rightarrow \mathcal{C}_h$		$\mathcal{S}_y \mathcal{C}_h \mathcal{F}_e \rightarrow \mathcal{O}_t$	
	Acc. \uparrow	$F_1 \uparrow$						
TextCNN (Kim, 2014)	74.98	85.67	74.61	79.58	80.95	88.54	69.78	68.43
EANN (Wang et al., 2018)	75.15	85.81	75.43	80.21	78.64	87.80	70.11	69.83
RoBERTa (Liu et al., 2019)	76.29	86.07	76.49	80.38	83.16	88.87	77.53	77.97
MDDA (Zhang et al., 2021)	76.10	85.92	76.81	80.54	82.42	88.21	79.51	78.68
Fish (Shi et al., 2022)	76.30	85.97	77.87	81.19	82.88	88.23	78.08	78.03
MetaAdapt (Yue et al., 2023)	76.81	86.19	78.78	81.59	83.74	88.99	78.65	78.32
DEAR (ours)	79.27	87.75	81.46	84.97	84.82	89.57	80.62	80.03

Table 3: Performance comparison between DEAR and other recent approaches in the *ITD* scenario using the PHEME corpus, which contains articles from four events with similar topics. The evaluation is conducted by combining data from three events and predicting the remaining one, which is unseen, for the fake news detection task. Our method achieves the highest accuracy and F_1 score, demonstrating its superior effectiveness in detecting fake news across different but related events.

Method	$\mathcal{P}_o \mathcal{G}_o \rightarrow \mathcal{C}_o$		$\mathcal{P}_o \mathcal{C}_o \rightarrow \mathcal{G}_o$		$\mathcal{C}_o \mathcal{G}_o \rightarrow \mathcal{P}_o$	
	Acc. \uparrow	$F_1 \uparrow$	Acc. \uparrow	$F_1 \uparrow$	Acc. \uparrow	$F_1 \uparrow$
TextCNN (Kim, 2014)	62.53	64.26	51.08	21.07	59.23	53.97
RoBERTa (Liu et al., 2019)	64.90	63.26	52.69	25.03	60.82	56.30
EANN (Wang et al., 2018)	63.45	64.02	50.64	23.99	62.15	61.43
MDDA (Zhang et al., 2021)	63.82	64.89	52.32	34.18	65.82	66.43
Fish (Shi et al., 2022)	63.98	65.01	54.58	37.74	66.59	68.49
MetaAdapt (Yue et al., 2023)	64.21	67.84	58.89	41.19	67.06	66.12
DEAR (ours)	66.34	68.93	63.43	48.97	70.04	69.80

Table 4: Performance comparison between DEAR and other recent approaches in the *CTD* scenario using three datasets corresponding to three distinct topics. The evaluation is conducted by combining data from two datasets and predicting the remaining one, which is unseen, for the fake news detection task. Our method achieves the highest accuracy and F_1 score, demonstrating its superior effectiveness in detecting fake news across events with different topics.

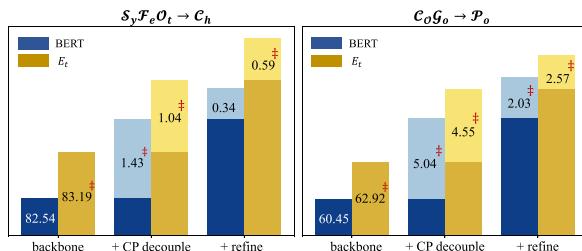


Figure 7: Results of ablation study, evaluated with accuracy performance on datasets from both *ITD* and *CTD* scenarios. Significance testing is indicated by \ddagger for $p < 0.005$.

4 Related Work

4.1 Fake News Detection

Content-aware fake news detection approaches are designed for detection analysis based on input claims. For example, pre-trained transformer models are leveraged to extract semantic or syntactic properties, enhancing their capability to detect

fake news (Ma et al., 2016; Chen et al., 2018; Das et al., 2021; Yue et al., 2022; Jiang et al., 2022; Li et al., 2024). Additionally, the integration of multi-modal input, combining text and image features, has been explored to further enhance detection performance (Santhosh et al., 2022; Shang et al., 2022b; Hu et al., 2024b).

Beyond content-based methods, there are approaches that leverage user interactions to assess the credibility of online posts (Jin et al., 2016). Likewise, analyzing patterns in propagation paths proves effective in detecting fake news on social media platforms (Shu et al., 2020). The incorporation of social attributes, such as user dynamics, enriches fake news detection by introducing contextual information (Shu et al., 2019; Nan et al., 2024). When integrated with a content-based module, fake news detection systems exhibit enhanced accuracy (Mosallanezhad et al., 2022; Lin et al., 2022).

There are also approaches that utilize external knowledge as augmentative features and support in the verification of facts and identifications of fake news (Brand et al., 2021). Approaches involving knowledge graphs or crowd-sourcing methodologies can be employed to extract supplementary information for fake news detection (Wu et al., 2024b; Shang et al., 2022a). But they usually need extra human annotations. More recently, the use of large language models (LLMs) has shown promising results in enhancing the performance of fake news detection (Hu et al., 2024a; Wu et al., 2024a).

Many current fake news detection approaches concentrate on news articles specific to in-event scenarios, which contain event-specific characteristics. This raises concerns about their effectiveness in unobserved events marked by event shifts. Our focus is channeled towards the systematic exploration of news articles from unseen events within the scenario of early fake news detection. This is particularly relevant in the early stage when the news has not been widely propagated and is primarily available as news content.

4.2 Domain Adaptation on Fake News Detection

Domain-adaptive fake news detection approaches aim to predict the news from unseen domains, addressing the challenges posed by domain shifts. Several approaches (Li et al., 2021; Yue et al., 2022; Lin et al., 2022; Silva et al., 2024) focus on domain adaptation, assuming access to a portion of target domain data during training. For instance, Silva et al. (2021) introduce an unsupervised technique for selecting unlabeled news records to maximize domain coverage and preserve both domain-specific and cross-domain knowledge through the disentanglement mechanism. Mosallanezhad et al. (2022) propose a domain adaptive detection framework using reinforcement learning and incorporating auxiliary information. Yue et al. (2023) propose a meta learning-based method for few-shot domain-adaptive misinformation detection, leveraging a few target examples to exploit source domain knowledge under the guidance of limited target data.

Incorporating cross-event scenarios into fake news detection has received less attention. Most approaches treat different events as distinct domains and use domain adaptation techniques to

tackle the event-generalized challenge. For instance, Wang et al. (2018) propose a multi-modal fake news detection framework using event adversarial networks, aiming to learn shared features across events by mitigating event-specific knowledge that is not shared among different events. Zhang et al. (2021) propose a disentangled domain adaptation mechanism for fake news detection, particularly for unseen events. Liu et al. (2024) argue that large-scale datasets might not generalize well to unseen events due to domain shifts and introduce inter-domain and cross-modality alignment modules that reduce domain shift and the modality gap.

However, these fake news detection methods operate within the domain adaptation framework, assuming access to some target domain data or correlated extra knowledge during training. This assumption can be problematic given the dynamic nature of fake news generalization and propagation, especially when target domain data is not accessible during the training phase.

5 Conclusion

In this paper, we introduce DEAR, an early fake news detection approach that leverages a disentanglement architecture to separate authenticity-related and event-specific knowledge. Our approach employs interactive cross-perturbation and refinement learning techniques to enhance the disentanglement effect, minimizing interactions between the decoupled representations. An adaptive multi-grained semantic encoder, based on BERT, generates hierarchical and fine-grained textual representations. Experimental results across multiple datasets demonstrate the effectiveness of DEAR in mitigating event-specific knowledge for fake news detection, outperforming state-of-the-art methods. As future work, we plan to extend the proposed disentangled methodology to address multi-modal fake news detection, exploring the possibility of mitigating event-specific knowledge in the multi-modality context.

Acknowledgments

We would like to thank Prof. Andrei Popescu-Belis for his insightful suggestions on revising the paper. We also extend our gratitude to the editor of TACL and the anonymous reviewers

for their valuable feedback. This work was supported partly by the National Natural Science Foundation of China (62402073 and 62172067), the National Social Science Foundation of China (24XMZ092), the Natural Science Foundation of Chongqing (CSTB2022NSCQ-MSX1342), and the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202300619).

References

Erik Brand, Kevin Roitero, Michael Soprano, and Gianluca Demartini. 2021. E-BART: Jointly predicting and explaining truthfulness. In *Proceedings of the 2021 Truth and Trust Online Conference*, pages 18–27.

Tong Chen, Xue Li, Hongzhi Yin, and Jun Zhang. 2018. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In *Proceedings of the Trends and Applications in Knowledge Discovery and Data Mining*, volume 11154, pages 40–52. https://doi.org/10.1007/978-3-030-04503-6_4

Sourya Dipta Das, Ayan Basak, and Saikat Dutta. 2021. A heuristic-driven ensemble framework for COVID-19 fake news detection. In *Proceedings of the Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pages 164–176. https://doi.org/10.1007/978-3-030-73696-5_16

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Jiangshu Du, Yingtong Dou, Congying Xia, Limeng Cui, Jing Ma, and Philip S. Yu. 2021. Cross-lingual COVID-19 fake news detection. In *Proceedings of the 2021 International Conference on Data Mining Workshops*, pages 859–862. <https://doi.org/10.1109/ICDMW53433.2021.00110>

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, volume 2 of *NIPS’14*, pages 2672–2680. <https://doi.org/10.1145/342262>

Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. 2024a. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, pages 22105–22113. <https://doi.org/10.1609/aaai.v38i20.30214>

WeiQi Hu, Ye Wang, Yan Jia, Qing Liao, and Bin Zhou. 2024b. A multi-modal prompt learning framework for early detection of fake news. In *Proceedings of the Eighteenth International AAAI Conference on Web and Social Media*, pages 651–662. <https://doi.org/10.1609/icwsm.v18i1.31341>

Xun Huang and Serge J. Belongie. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the European Conference on Computer Vision*, pages 1510–1519. <https://doi.org/10.1109/ICCV.2017.167>

Gongyao Jiang, Shuang Liu, Yu Zhao, Yueheng Sun, and Meishan Zhang. 2022. Fake news detection via knowledgeable prompt learning. *Information Processing & Management*, 59(5):103029. <https://doi.org/10.1016/J.IPM.2022.103029>

Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 2972–2978. <https://doi.org/10.1609/aaai.v30i1.10382>

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751. <https://doi.org/10.3115/v1/D14-1181>

Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3402–3413.

Jiayang Li, Xuan Feng, Tianlong Gu, and Liang Chang. 2024. Dual-teacher de-biasing distillation framework for multi-domain fake news detection. In *Proceedings of the 40th IEEE International Conference on Data Engineering*, pages 3627–3639. <https://doi.org/10.1109/ICDE60146.2024.00279>

Yichuan Li, Kyumin Lee, Nima Kordzadeh, Brenton D. Faber, Cameron Fiddes, Elaine Chen, and Kai Shu. 2021. Multi-source domain adaptation with weak supervision for early fake news detection. In *Proceedings of the 2021 IEEE International Conference on Big Data*, pages 668–676. <https://doi.org/10.1109/BigData52589.2021.9671592>

Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Guang Chen. 2022. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. In *Proceedings of the Findings of the Association for Computational Linguistics*, pages 2543–2556. <https://doi.org/10.18653/v1/2022.findings-naacl.194>

Hui Liu, Wenya Wang, Hao Sun, Anderson Rocha, and Haoliang Li. 2024. Robust domain misinformation detection via multi-modal feature alignment. *IEEE Transactions on Information Forensics and Security*, 19:793–806. <https://doi.org/10.1109/TIFS.2023.3326368>

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 3818–3824.

Ahmadreza Mosallanezhad, Mansooreh Karami, Kai Shu, Michelle V. Mancenido, and Huan Liu. 2022. Domain adaptive fake news detection via reinforcement learning. In *Proceedings of the ACM Web Conference 2022*, pages 3632–3640. <https://doi.org/10.1145/3485447.3512258>

Qiong Nan, Qiang Sheng, Juan Cao, Beizhe Hu, Danding Wang, and Jintao Li. 2024. Let silence speak: Enhancing fake news detection with generated comments from large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 1732–1742. <https://doi.org/10.1145/3627673.3679519>

Nikita Mariam Santhosh, Jo Cherian, and Lekshmi S. Nair. 2022. A multi-model intelligent approach for rumor detection in social networks. In *Proceedings of the 2022 International Conference on Computing, Communication, Security and Intelligent Systems*, pages 1–5. <https://doi.org/10.1109/IC3SIS54991.2022.9885398>

Lanyu Shang, Ziyi Kou, Yang Zhang, Jin Chen, and Dong Wang. 2022a. A privacy-aware distributed knowledge graph approach to qos-driven COVID-19 misinformation detection. In *Proceedings of the 30th IEEE/ACM International Symposium on Quality of Service*, pages 1–10. <https://doi.org/10.1109/IWQoS54832.2022.9812879>

Lanyu Shang, Ziyi Kou, Yang Zhang, and Dong Wang. 2022b. A duo-generative approach to explainable multimodal COVID-19 misinformation detection. In *Proceedings of the ACM Web Conference 2022*, pages 3623–3631. <https://doi.org/10.1145/3485447.3512257>

Yuge Shi, Jeffrey Seely, Philip H. S. Torr, Siddharth Narayanaswamy, Awni Y. Hannun, Nicolas Usunier, and Gabriel Synnaeve. 2022. Gradient matching for domain generalization. In *Proceedings of the Tenth International Conference on Learning Representations*.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. Fakenewsnets: A data repository with news content, social context and dynamic information for studying fake news on social media. *Big Data*, abs/1809.01286.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, and Huan Liu. 2020. Hierarchical propagation networks for fake news detection: Investigation and exploitation. In *Proceedings*

of the Fourteenth International AAAI Conference on Web and Social Media, pages 626–637. <https://doi.org/10.1609/icwsm.v14i1.7329>

Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 312–320. <https://doi.org/10.1145/3289600.3290994>

Amila Silva, Ling Luo, Shanika Karunasekera, and Christopher Leckie. 2021. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. In *Proceedings of the AAAI conference on artificial intelligence*, pages 557–565. <https://doi.org/10.1609/aaai.v35i1.16134>

Amila Silva, Ling Luo, Shanika Karunasekera, and Christopher Leckie. 2024. Unsupervised domain-agnostic fake news detection using multi-modal weak signals. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):7283–7295. <https://doi.org/10.1109/TKDE.2024.3392788>

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. 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*, pages 849–857. <https://doi.org/10.1145/3219819.3219903>

Jiaying Wu, Jiafeng Guo, and Bryan Hooi. 2024a. Fake news in sheep’s clothing: Robust fake news detection against llm-empowered style attacks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3367–3378. <https://doi.org/10.1145/3637528.3671977>

Junfei Wu, Weizhi Xu, Qiang Liu, Shu Wu, and Liang Wang. 2024b. Adversarial contrastive learning for evidence-aware fake news detection with graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):5591–5604. <https://doi.org/10.1109/TKDE.2023.3341640>

Zhenrui Yue, Huimin Zeng, Ziyi Kou, Lanyu Shang, and Dong Wang. 2022. Contrastive domain adaptation for early misinformation detection: A case study on COVID-19. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2423–2433. <https://doi.org/10.1145/3511808.3557263>

Zhenrui Yue, Huimin Zeng, Yang Zhang, Lanyu Shang, and Dong Wang. 2023. Metaadapt: Domain adaptive few-shot misinformation detection via meta learning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 5223–5239. <https://doi.org/10.18653/v1/2023.acl-long.286>

Huaiwen Zhang, Shengsheng Qian, Quan Fang, and Changsheng Xu. 2021. Multimodal disentangled domain adaption for social media event rumor detection. *IEEE Transactions on Multimedia*, 23:4441–4454. <https://doi.org/10.1109/TMM.2020.3042055>