Supplementary Materials

Anonymous submission

Related Work

Multi-Domain Fake News Detection

In the wake of social media's ascent, a proliferation of news across various domains has emerged. Multi-domain fake news detection (Fu, Peng, and Liu 2023; Wu et al. 2024) refers to the techniques and methodologies for identifying and detecting fake news across multiple distinct journalistic domains. To tackle the challenges of distributional disparities and domain shifts in multi-domain data, MDFEND (Nan et al. 2021) incorporates a domain gate and an expert network to extract semantic information specific to news within distinct domains. M^3FEND (Zhu et al. 2022) introduces the concept of a domain memory bank designed to automatically capture and store domain features, identifying potential domain labels, thereby addressing the issue of incomplete domain labeling through a multi-view approach. However, the generalizability of these models across different domains may be limited, particularly for those domains that are not adequately represented in the training data. (Silva et al. 2021) proposed a new framework that is capable of preserving both domain-specific and cross-domain knowledge within news records. By constructing a knowledge graph and incorporating entity triples, KG-MFEND (Chen, Fu, and Tang 2023) develops sentence trees to capture the specific semantics of news across different domains. (LI, SANG, and ZHANG 2024) proposes a multi-domain fake news detection model, Transm3, which is enhanced by the Adaptive Pooling Kernel Convolutional Neural Network and Transformer. The model utilizes a multi-granularity cross-domain interactor for feature view combination. StruACL (Wei et al. 2025) contrasts representations between content-only data and propagation-rich data, then, a propagation-guided adversarial training strategy is designed to enhance the diversity of representations. HFGD (Wang et al. 2025) explicitly leverages a heterogeneous network to model multiple relationships. This method can mitigate domain bias and comprehensively capture news diversity and inter-domain interactions. Specifically, domain cohesion grounded in news semantics is devised to capture the relevance of news within a domain. However, these approaches neglect the transferable information across domains and are confined to a single modality, resulting in the loss of much crucial information during extraction.

Multimodal Fake News Detection

Multimodal fake news detection is a method that utilizes various types of data, such as text, images, videos, audio, etc., to identify and detect fake news. MVAE (Khattar et al. 2019) proposes a novel multimodal variational autoencoder model capable of simultaneously processing multiple data sources, including text, images, and propagation networks, utilizing variational inference to optimize the latent representation. MCNN (Xue et al. 2021) has designed a comprehensive framework that integrates feature extraction, consistency scoring, and classifiers to achieve automated fake news detection. It employs a consistency scoring mechanism to measure the degree of matching between features from different modalities. (Chen et al. 2022) emphasizes the importance of considering cross-modal ambiguity in multimodal fake news detection and designs a novel framework that enhances detection accuracy through learning from cross-modal ambiguity. MMDFND (Tong et al. 2024) extracts domain-specific semantics through a progressive hierarchical extraction network that includes domain embedding and attention mechanisms, and utilizes a pivot transformer network to integrate information from different modalities. SSA-MFND (Shang et al. 2025) compares the semantic consistency between the description layer and the entity layer, projecting the semantic representation into isotropic space before the semantic consistency calculation. KAMP (Zhang et al. 2025) encompasses an innovative multimodal learning architecture alongside a set of elaborate pre-training tasks, engineered to concurrently extract meaningful knowledge from unimodal data, multimodal interactions, and background knowledge graphs. Gamed (Shen et al. 2025) employs multiple parallel expert networks for feature refinement and incorporates pre-embedded semantic knowledge to strengthen the experts' capabilities in perspective sharing. The aforementioned methods overlook the complementarity and differences between different modalities during information integration. The simple fusion of modal features can lead to negative transfer on the model's performance.

Datasets

Our model is evaluated on three real-world datasets: two Chinese multi-domain datasets, Weibo (Wang et al. 2018) and Weibo-21 (Nan et al. 2021), and one English multi-domain dataset, Fine-Fake (Zhou et al. 2024). The Weibo and Weibo-21 datasets are divided into nine domains, while FineFake contains six domains and one uncategorized domain.

Table 1: Data Statistics of Weibo

domain	Science	Military	Education	International	Politics
real	139	160	273	45	112
fake	91	122	148	41	164
all	230	282	421	86	276
domain	**Health**	**Finance**	**Entertainment**	**Society**	**All**
real	186	137	1,120	1,443	3,615
fake	519	143	508	2,372	4,108
all	705	280	1,628	3,815	7,723

Experiment Settings

We implement our model using PyTorch (Paszke et al. 2019) and conduct all experiments on NVIDIA RTX 4090 GPUs. For text processing, we use a pre-trained BERT (Devlin et al. 2019) model, setting the maximum sequence length to 197. For image processing, we use a pre-trained ViT (Dosovitskiy et al. 2020) model with input images resized to 224×224 pixels. For the initial multimodal representations, we utilize CN-CLIP (Yang et al. 2022) for the Chinese datasets and CLIP (Radford et al. 2021) for the English dataset. In our Domain-gated Expert Network (DExNet), the number of experts L_{exp} is set to 8. For the Domain-aware Modal Prompt Generator (DMPG), we use $L_p = 4$ prompts per modality view. The number of domain prototypes M for the PAD metric is set to 16. The Gumbel-based Neighbor Selector (GNS) is configured to select the top-$S = 2$ neighbor domains. We set the total number of training epochs to 50. To prevent overfitting and ensure optimal convergence, an early stopping strategy is employed: training is terminated if the validation performance (monitored by F1-score) does not improve for 3 consecutive epochs, and the model parameters corresponding to the best validation performance are retained. We train the model using the AdamW optimizer with a learning rate of 1e-4 and a batch size of 32. The balancing hyperparameter λ for the reconstruction loss is set to 0.1. Performance is evaluated using Accuracy, F1-score, and AUC.

Table 2: Data Statistics of Weibo21

domain	Science	Military	Education	Disasters	Politics
real	143	121	243	185	306
fake	93	222	248	591	546
all	236	343	491	776	852
domain	**Health**	**Finance**	**Entertainment**	**Society**	**All**
real	485	959	1,000	1,198	4,640
fake	515	362	440	1,471	4,488
all	1,000	1,321	1,440	2,669	9,128

Table 3: Data Statistics of FineFake

domain	Politics	Entertainment	Business	Health	Society
real	3,722	2,514	527	438	2,236
fake	2,005	1,185	476	272	1,703
all	5,727	3,699	1,003	710	3,939
domain	**Conflict**	**Stream**	**Official**	**Fact-Check**	**All**
real	979	3,895	4,138	2,474	10,507
fake	739	1,105	215	5,082	6,402
all	1,718	5,000	4,353	7,556	16,909

Baselines

To comprehensively assess the performance of our proposed model, we compare it against baseline methods organized into three categories: (1) single-modal multi- domain methods, including: MMoE (Ma et al. 2018), MoSE (Qin et al. 2020), MDFEND (Nan et al. 2021), M^{3}DFEND (Zhu et al. 2022) and PLDFEND (Peng et al. 2023); (2) multimodal multi-domain methods, including: KATMF (Song et al. 2021), MMDFND (Tong et al. 2024) and DAMMFND (Lu, Tong, and Ye 2025); Descriptions of baseline methods are as follows:

1. **Single Modal Multi-Domain Methods**:
   - **MMoE**, which utilizes a mixture of experts (MoE) for multi-domain fake news detection.
   - **MoSE**, which replaces the experts in the MMoE framework with Long Short-Term Memory networks (LSTM).
   - **MDFEND**, which uses domain gating to perform weighted aggregation of information from experts.
   - **M^{3}DFEND**, which substitutes the experts in MDFEND with text semantic, emotion, and style extractor, while replacing domain gate with domain adapter.
   - **PLDFEND**, which places particular emphasis on refining prompt integration and classification logic within the model's framework.
2. **Multimodal Multi-Domain Methods**:
   - **KATMF**, which employs adversarial multi-task learning and knowledge-enhanced Transformer to model the differences in feature distributions across various domains.
   - **MMDFND**, which utilizes an improved PLE module to capture both the commonalities and specificities among domains, and incorporates AdaIN for multimodal fusion.
   - **DAMMFND**, which extracts more accurate domain information through Domain Disentanglement, while simultaneously mitigating negative transfer between domains.

References

Chen, Y.; Li, D.; Zhang, P.; Sui, J.; Lv, Q.; Tun, L.; and Shang, L. 2022. Cross-modal ambiguity learning for multimodal fake news detection. In *Proceedings of the ACM web conference 2022*, 2897–2905.

Chen, Z.; Fu, C.; and Tang, X. 2023. Multi-domain Fake News Detection with Fuzzy Labels. In *International Conference on Database Systems for Advanced Applications*, 331–343. Springer.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 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, volume 1 (long and short papers)*, 4171–4186.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Fu, L.; Peng, H.; and Liu, S. 2023. KG-MFEND: an efficient knowledge graph-based model for multi-domain fake news detection. *The Journal of Supercomputing*, 79(16): 18417–18444.

Khattar, D.; Goud, J. S.; Gupta, M.; and Varma, V. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *The world wide web conference*, 2915–2921.

LI, J.; SANG, G.; and ZHANG, Y. 2024. APK-CNN and Transformer-enhanced multi-domain fake news detection model. *Journal of Computer Applications*, 0.

Lu, W.; Tong, Y.; and Ye, Z. 2025. DAMMFND: Domain-Aware Multimodal Multi-view Fake News Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 559–567.

Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; and Chi, E. H. 2018. 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*, 1930–1939.

Nan, Q.; Cao, J.; Zhu, Y.; Wang, Y.; and Li, J. 2021. MDFEND: Multi-domain fake news detection. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 3343–3347.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Peng, Z.; Lin, N.; Zhou, Y.; Zhou, D.; and Yang, A. 2023. Prompt learning for low-resource multi-domain fake news detection. In *2023 international conference on asian Language Processing (IALP)*, 314–319. IEEE.

Qin, Z.; Cheng, Y.; Zhao, Z.; Chen, Z.; Metzler, D.; and Qin, J. 2020. Multitask mixture of sequential experts for user activity streams. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 3083–3091.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.

Shang, W.; Song, K.; Ji, J.; Yi, T.; Cai, J.; and Li, X. 2025. Semantic space aligned multimodal fake news detection. *Information Fusion*, 103469.

Shen, L.; Long, Y.; Cai, X.; Razzak, I.; Chen, G.; Liu, K.; and Jameel, S. 2025. Gamed: Knowledge adaptive multi-experts decoupling for multimodal fake news detection. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, 586–595.

Silva, A.; Luo, L.; Karunasekera, S.; and Leckie, C. 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*, volume 35, 557–565.

Song, C.; Ning, N.; Zhang, Y.; and Wu, B. 2021. Knowledge augmented transformer for adversarial multidomain multi-classification multimodal fake news detection. *Neurocomputing*, 462: 88–100.

Tong, Y.; Lu, W.; Zhao, Z.; Lai, S.; and Shi, T. 2024. MMDFND: Multi-modal multi-domain fake news detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 1178–1186.

Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 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*, 849–857.

Wang, Y.; Yuan, S.; Li, W.; Feng, Y.; Yu, X.; Liu, F.; Wang, C.; and Pan, Q. 2025. Heterogeneous network for Hierarchical Fine-Grained Domain Fake News Detection. *Information Processing & Management*, 62(4): 104141.

Wei, L.; Hu, D.; Zhou, W.; Yu, P. S.; and Hu, S. 2025. Structure-adaptive Adversarial Contrastive Learning for Multi-Domain Fake News Detection. In *Findings of the Association for Computational Linguistics: ACL 2025*, 9739–9752.

Wu, D.; Tan, Z.; Zhao, H.; Jiang, T.; and Qi, M. 2024. LIMFA: label-irrelevant multi-domain feature alignment-based fake news detection for unseen domain. *Neural Computing and Applications*, 36(10): 5197–5215.

Xue, J.; Wang, Y.; Tian, Y.; Li, Y.; Shi, L.; and Wei, L. 2021. Detecting fake news by exploring the consistency of multimodal data. *Information Processing & Management*, 58(5): 102610.

Yang, A.; Pan, J.; Lin, J.; Men, R.; Zhang, Y.; Zhou, J.; and Zhou, C. 2022. Chinese CLIP: Contrastive Vision-Language Pretraining in Chinese. *arXiv preprint arXiv:2211.01335*.

Zhang, L.; Zhang, X.; Zhou, Z.; Zhang, X.; Yu, P. S.; and Li, C. 2025. Knowledge-aware multimodal pre-training for fake news detection. *Information Fusion*, 114: 102715.

Zhou, Z.; Zhang, X.; Zhang, L.; Liu, J.; Wang, S.; Liu, Z.; Zhang, X.; Li, C.; and Yu, P. S. 2024. FineFake: A Knowledge-Enriched Dataset for Fine-Grained Multi-Domain Fake News Detection. *arXiv preprint arXiv:2404.01336*.

Zhu, Y.; Sheng, Q.; Cao, J.; Nan, Q.; Shu, K.; Wu, M.; Wang, J.; and Zhuang, F. 2022. Memory-guided multi-view multi-domain fake news detection. *IEEE Transactions on Knowledge and Data Engineering*.