COOL: Comprehensive Knowledge Enhanced Prompt Learning for Domain Adaptive Few-shot Fake News Detection

Anonymous ACL submission

Abstract

Most Fake News Detection (FND) methods often struggle with data scarcity for emerging news domain. Recently, prompt learning based on Pre-trained Language Models (PLM) has emerged as a promising approach in domain adaptive few-shot learning, since it greatly reduces the need for labeled data by bridging the gap between pre-training and downstream task. Furthermore, external knowledge is also helpful in verifying emerging news, as emerging news often involves timely knowledge that may not be contained in the PLM's outdated prior knowledge. To this end, we propose COOL, a Comprehensive knOwledge enhanced prOmpt Learning method for domain adaptive few-shot FND. Specifically, we propose a comprehensive knowledge extraction module to extract both structured and unstructured knowledge that are positively or negatively correlated with news from external sources, and adopt an adversarial contrastive enhanced hybrid prompt learning strategy to model the domain-invariant news-knowledge interaction pattern for FND. Experimental results demonstrate the superiority of COOL over various state-of-the-arts.

1 Introduction

Emerging news domain with limited labeled data often have distinctive semantic characteristics other than historical news domain with sufficient labeled data, leading to degenerated performance for PLM-based FND methods which have to be fine-tuned on large-scale labeled data. To improve FND on emerging target domain, various domain adaptive fine-tuning strategies on PLM have been investigated (Mehta et al., 2022; Kaliyar et al., 2021; Li et al., 2023; Mridha et al., 2021). However, fine-tuning PLM is inherently data-intensive, as it requires additional supervised signals to adapt PLM from pre-training task to downstream task. Recently, prompt learning which bridges the gap between pre-training and downstream task by keeping

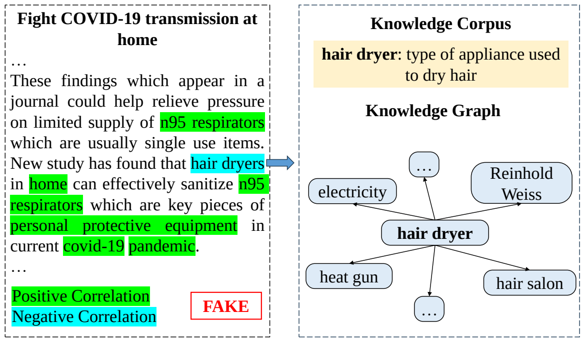


Figure 1: A piece of real-world fake news about the prevention of Covid-19.

downstream learning the same as the pre-training process, achieves success in few-shot scenarios and has been used in various domain adaptive tasks (Bai et al., 2024; Guo et al., 2024; Ge et al., 2023).

Despite its promising performance, PLM's prior knowledge is constrained by the outdated pre-training corpus, leading to sub-optimal detection performance in emerging news domain where timely and domain-specific knowledge is involved. Therefore, it is crucial to leverage up-to-date heterogeneous external knowledge, including structured knowledge graph with relational knowledge among entities and unstructured knowledge corpus with descriptions about entity properties (Speer et al., 2017; Pei et al., 2023), to assist in domain adaptive FND. Most previous studies generally extract entity knowledge positively correlated with news (Dun et al., 2021; Tseng et al., 2022; Ma et al., 2023). However, the negatively correlated knowledge, i.e., entity knowledge not very correlated with news semantics, may also contributes considerably to FND. For example, Figure 1 illustrates a fake news concerning the prevention of Covid-19, where exists entities positively correlated with news semantics like "pandemic", as well as a negatively correlated entity "hair dryer". Intuitively, the negatively correlated entity "hair dryer", whose knowledge greatly deviates from the news semantics, significantly reveals the authenticity of the news. Hence both positively and negatively correlated knowledge in heterogeneous source should be comprehensively extracted for domain adaptive few-shot FND.

Existing knowledge enhanced FND models typically inject knowledge by concatenating it with the learned news features before the final classifier and adopting much labeled data to capture their interaction patterns (Ma et al., 2023; Dun et al., 2021). Such scheme may not be applicable to few-shot prompt learning, as its input feature for classifier is the learned embedding of a mask word where the knowledge-news interaction patterns are hard to be captured. Another intuitive scheme is to incorporate the knowledge into the prompt template before PLM encoder. However, both hand-crafted and soft prompt template may not be suitable for directly injecting knowledge, as hard template cannot flexibly inject various forms and quantities of knowledge, while soft template may struggle to fit the FND task. As a result, hybrid templates have been adopted, which incorporate knowledge representation into several soft prompt vectors, while guide PLM in reasoning about news authenticity via hard templates (Jiang et al., 2022). Despite their effectiveness in modeling relationships between news, knowledge and detection task, their performances can be further improved in domain adaptive few-shot scenario by capturing the domain-invariant interaction patterns.

To this end, we propose COOL for domain adaptive few-shot FND, which extracts comprehensive knowledge that are positively or negatively correlated with news from heterogeneous sources, and injects it into prompt learning by an adversarial contrastive enhanced hybrid prompt learning framework. More specifically, a comprehensive knowledge extraction module is proposed to retrieve both structured relational and unstructured descriptive knowledge from external sources, and filter both positively and negatively correlated knowledge via a signed correlation-aware attention. The filtered comprehensive knowledge is incorporated by a hybrid prompt learning framework, where prefix soft prompt composed of several learnable tokens is used to receive knowledge representations flexibly, while postfix hand-crafted hard prompt facilitates PLM modeling task-specific interaction between knowledge and news. The adversarial contrastive learning is applied to facilitate the model capturing domain-invariant news-knowledge interaction patterns to improve the domain adaptive few-shot FND performance. Experimental studies validate the benefits of incorporating comprehensive knowledge into prompt learning for domain adaptive few-shot FND. The primary contributions of this paper can be summarized as:

(1) We highlight that the comprehensive knowledge positively or negatively correlated with news is crucial for PLM to detect fake news in emerging domains, which can be extracted from heterogeneous source.

(2) We propose COOL, which devises a comprehensive knowledge extraction module to extract knowledge and injects it into a hybrid prompt learning framework to model domain-invariant news-knowledge interaction patterns.

(3) Experiments on real-word datasets are conducted to demonstrate that COOL consistently outperforms the several state-of-the-arts.

2 Related Works

Domain Adaptive Few-shot News Detection. Previous FND methods focus on modeling fake news patterns by PLM-based models fine-tuned on large-scale labeled datasets (Mehta et al., 2022; Kaliyar et al., 2021; Xiao et al., 2024; Mridha et al., 2021). However, it is frequent to face the data scarcity issue of emerging news domain. To tackle this problem, many domain adaptive few-shot FND methods have adopted various techniques to adapt the domain-invariant features learned from the abundant source domain data to the target news domain with limited labeled data (Yue et al., 2022; Mosallanezhad et al., 2022; Lin et al., 2022; Ran and Jia, 2023), such as meta-learning improving domain adaptation by adjusting model parameters step by step across tasks (Yue et al., 2023; Nan et al., 2022; Hospedales et al., 2021), and contrastive learning reducing the inter-domain discrepancy by appropriate contrastive loss (Yue et al., 2022; Lin et al., 2022; Ran and Jia, 2023). More recently, prompt learning, which bridges the gap between PLM's pre-training and downstream task, exhibits significant successes in many domain adaptive few-shot tasks, such as rumor detection (Lin et al., 2023) and dialogue summarization (Zhao et al., 2022). Despite its success in various tasks, its performance may be constrained in domain adaptive FND, as the emerging news typically involves timely and domain-specific knowledge that may not be included in PLM's outdated pre-training corpus. This inspires us to design a knowledge enhanced prompt learning method for better domain adaptive few-shot FND.

Knowledge Enhanced Fake News Detection. News naturally encompasses a number of knowledge entities, whose knowledge can serve as critical evidence for news verification, inspiring researchers to investigate knowledge-enhanced methods for FND (Dun et al., 2021; Tseng et al., 2022; Ma et al., 2023; Zhang et al., 2024). Most existing methods leverage structured knowledge graph, e.g. ConceptNet (Speer et al., 2017) and YAGO (Suchanek et al., 2007), to capture the relational knowledge among entities for FND (Ma et al., 2023; Kim et al., 2023; Sun et al., 2023). Some studies benefit from abundant heterogeneous knowledge by exploiting both structured knowledge graph and unstructured knowledge corpus, e.g., Wikipedia corpus (Hu et al., 2021; Zou et al., 2023). Timely and rich external knowledge can compensate for the knowledge gap of PLM in emerging news, thereby improving the domain adaptive few-shot FND performance of PLM based methods, including prompt learning. KPL (Jiang et al., 2022) devises a knowledgeable prompt learning framework which incorporates the sequential knowledge entities into prompt template to predict the news veracity. Different from Jiang et al. (2022), the proposed COOL extracts more comprehensive knowledge that have either positive or negative correlation with news from both structured and unstructured external knowledge sources, which is further incorporated into an adversarial contrastive enhanced hybrid prompt learning framework to model the domain-invariant interaction patterns between news and knowledge for domain adaptive few-shot FND.

3 Problem Statement

Let $\mathcal{D}_s = \{(\mathcal{X}_1^s, y_1^s), (\mathcal{X}_2^s, y_2^s), \ldots, (x_M^s, y_M^s)\}$ and $\mathcal{D}_t = \{(\mathcal{X}_1^t, y_1^t), (\mathcal{X}_2^t, y_2^t), \ldots, (x_N^t, y_N^t)\}$ denote the datasets of source and target domain, respectively, where M, N denote the numbers of news in source and target domain, respectively. Each news $\mathcal{X} = \{w_i\}$ consists of a sequence of words. The label $y \in [0, 1]$ denotes the veracity of news, where 0 indicates true and 1 indicates fake. The domain adaptive few-shot FND is defined as: given the source domain dataset $\mathcal{D}_s$ and limited access to the target domain dataset, i.e., only a K-shot subset $\mathcal{D}_t' \subset \mathcal{D}_t$ is available for training where $K \ll N$, the goal is to correctly predict the veracity of news in the target domain dataset $\mathcal{D}_t$.

4 Methodologies

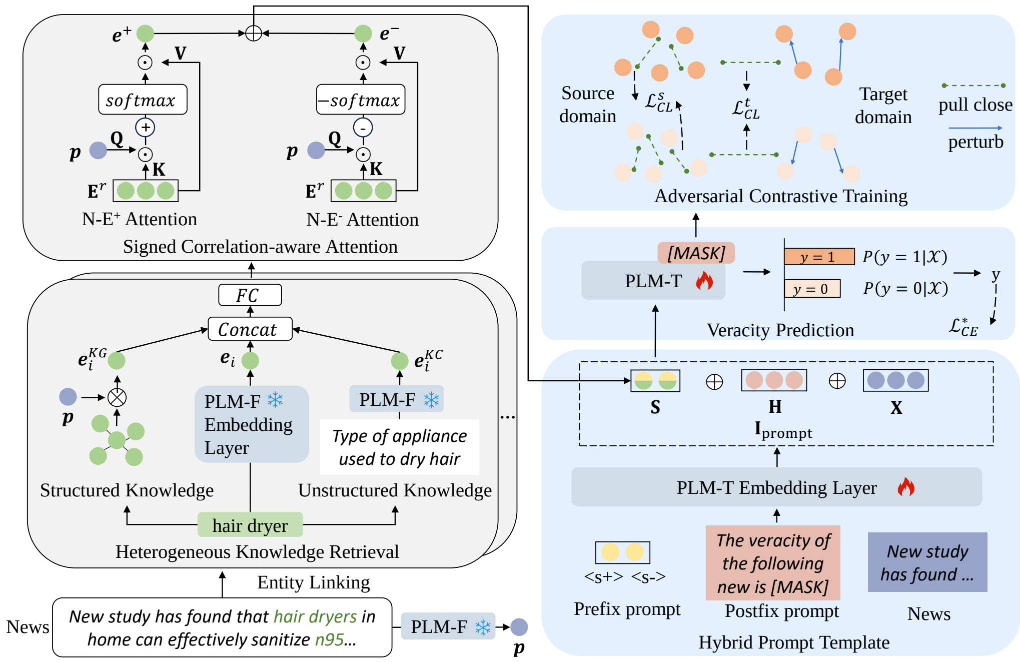
The architecture of the proposed model is illustrated in Figure 2. It consists of two modules: (i) comprehensive knowledge extraction, which extracts structured and unstructured knowledge from external sources that are positively or negatively correlated with news; (ii) hybrid prompt learning, in which a hybrid prompt template is devised to simultaneously incorporate external knowledge and guide PLM on FND task, and an adversarial contrastive training strategy is leveraged to capture the domain-invariant news-knowledge interaction pattern. Each of module is described in details next.

4.1 Comprehensive Knowledge Extraction

4.1.1 Heterogeneous Knowledge Retrieval

To retrieve external knowledge critical for verifying news, we first identify the knowledge entities $\mathcal{E} = \{en_i\}$ from a given news $\mathcal{X}$ by an entity linking method (Ferragina and Scaiella, 2010). The identified entities are embedded by the embedding layer of a parameter-frozen PLM, i.e., $\mathbf{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{|\mathcal{E}|}] \in \mathbb{R}^{|\mathcal{E}| \times d}$, where $\boldsymbol{e}_i$ is the averaged word embedding for entity en_i and d is the hidden dimension of PLM. The parameter-frozen PLM, denoted as PLM-F, is used to represent various information and will not attend training. The knowledge entities usually associate with heterogeneous external knowledge, including structured relational knowledge from knowledge graph, and unstructured descriptive knowledge from knowledge corpus. Therefore, we propose a structured knowledge retriever and an unstructured knowledge retriever for heterogeneous knowledge.

Structured Knowledge Retriever. Knowledge graph $\mathcal{G}$ encompasses structured knowledge in form of triples (en_s, rel, en_t), where rel is the relation between two entities. The structured knowledge of an entity en_i is set as its neighbors $\mathcal{N}(en_i) = \{en \mid (en_i, rel, en) \in \mathcal{G} \vee (en, rel, en_i) \in \mathcal{G}\}$. Since a news entity may have many neighbors with various semantics, not all of its neighbors contribute equally for verifying a news. For instance, when verifying a news reporting Donald Trump signed the Iran Deal, the structured knowledge *(Donald Trump, significant event, United States withdrawal from Iran Deal)* is more informative than other knowledge of entity "Donald Trump".



(i) Comprehensive Knowledge Extraction **(ii) Hybrid prompt Learning**

Figure 2: Architecture of the proposed COOL model.

Hence the structured knowledge of each entity should be filtered based on their relevance with news semantics. The typical relevant information filter method is attention mechanism (Vaswani et al., 2017). However, since attention mechanism will also be used to filter relevant entities in news latter, if it is used simultaneously to filter the structured knowledge of each entity, the nested attention will be formed and lead to exponential increase in computational complexity. Previous works abandon the attentive filtering of structured knowledge by leveraging *mean pooling* to avoid the nested attention(Dun et al., 2021; Tseng et al., 2022). Instead, we adopt a modulation mechanism based on *product & max pooling* (Ouyang et al., 2022) to attentively filter the related structured knowledge with lower computational complexity. Specifically, given an entity en_i and its neighbors $\mathcal{N}(en_i)$, the structured knowledge is filtered as:

$$\boldsymbol{e}_i^{KG} = MP_{en \in \mathcal{N}(en_i)} \left(\boldsymbol{p} \otimes \boldsymbol{e}_{en} \right) \tag{1}$$

where $\boldsymbol{p}$ is the news representation embedded by PLM-F on its content, MP indicates the max pooling operation, $\otimes$ stands for the element-wise product, and $\boldsymbol{e}_{en}$ is the embedding of a neighbor entity. The element-wise product evaluates the relatedness between each neighbor and news, while the max pooling helps to focus on the most related knowledge from neighbors and reduce noises.

Unstructured Knowledge Retriever. The unstructured descriptive knowledge from knowledge corpus is another important supplement for FND, since it describes connotation and properties of each entity with natural language, whose semantics can also interact with news. For example, a short description "contagious disease caused by SARS-CoV-2" of entity "Covid-19" can interact with news by offering PLM with knowledge lacked during pre-training. The unstructured knowledge $\boldsymbol{e}_i^{KC} \in \mathbb{R}^d$ of a given entity en_i can also be embedded by applying PLM-F on its description sentence.

Heterogeneous knowledge complements each other for FND. The final knowledge $\boldsymbol{e}_i^r \in \mathbb{R}^d$ of an entity en_i is extracted by concatenating its structured knowledge $\boldsymbol{e}_i^{KG}$, initial entity embedding $\boldsymbol{e}_i$ and unstructured knowledge $\boldsymbol{e}_i^{KC}$ and passing through a fully connected layer FC:

$$\boldsymbol{e}_i^r = FC\left(\left[\boldsymbol{e}_i^{KG}; \boldsymbol{e}_i; \boldsymbol{e}_i^{KC}\right]\right) \tag{2}$$

4.1.2 Signed Correlation-aware Attention

The knowledge either positively or negatively correlated with news is significant for FND, as the positive one provides news-related knowledge context, while the negative one reveals news-knowledge discrepancy (Dun et al., 2021; Sun et al., 2023). To capture both positively and negatively correlated knowledge, a signed correlation-aware attention consisting of *News towards Positively correlated*

Entity (N-E$^+$) attention and *News towards Negatively correlated Entity* (N-E$^-$) attention is devised.

N-E$^+$ Attention. N-E$^+$ attention follows typical attention mechanism which captures positively correlated knowledge by assigning greater importance for entity knowledge that is more correlated with news semantics:

$$Attn^{+}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \operatorname{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{T}}{\sqrt{d_K}}\right)\mathbf{V} \quad (3)$$

where d_K is the dimension of keys. The dot-product attention function measures the positive correlation between queries and keys and assigns weights attentively to values. Hence the positively correlated knowledge $\boldsymbol{e}^+$ can be extracted by setting the news representation $\boldsymbol{p}$ as queries, and the all extracted knowledge $\mathbf{E}^r = \left[\boldsymbol{e}_1^r, \ldots, \boldsymbol{e}_{|\mathcal{E}|}^r\right] \in \mathbb{R}^{|\mathcal{E}| \times d}$ as keys and values:

$$\boldsymbol{e}^+ = Attn^+\left(\boldsymbol{p}\mathbf{W}_Q^+, \mathbf{E}^r\mathbf{W}_K^+, \mathbf{E}^r\mathbf{W}_V^+\right) \quad (4)$$

where $\mathbf{W}_Q^+$, $\mathbf{W}_K^+$, $\mathbf{W}_V^+$ are learnable parameters.

N-E$^-$ Attention. (N-E$^-$) attention captures negatively correlated knowledge by assigning greater importance for entity knowledge that is less correlated with news semantics:

$$Attn^{-}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = -\operatorname{softmax}\left(-\frac{\mathbf{Q}\mathbf{K}^{T}}{\sqrt{d_K}}\right)\mathbf{V} \quad (5)$$

The minus inside the dot-product attention function assigns greater weights to keys that are less correlated with queries, and the outside minus further reverses the direction of the resulted feature vector to distance it from the result of N-E$^+$ attention. Similarly, the negatively correlated knowledge $\boldsymbol{e}^-$ is extracted as:

$$\boldsymbol{e}^- = Attn^-\left(\boldsymbol{p}\mathbf{W}_Q^-, \mathbf{E}^r\mathbf{W}_K^-, \mathbf{E}^r\mathbf{W}_V^-\right) \quad (6)$$

where $\mathbf{W}_Q^-$, $\mathbf{W}_K^-$, $\mathbf{W}_V^-$ are learnable parameters.

Finally, we obtain the comprehensive knowledge as $\mathbf{E}^c = [\boldsymbol{e}^+, \boldsymbol{e}^-] \in \mathbb{R}^{2 \times d}$ for FND.

4.2 Hybrid Prompt Learning

4.2.1 Hybrid Prompt Template

To inject comprehensive knowledge into prompt learning, a hybrid prompt template consists of both prefix soft prompt and postfix hard prompt is adopted. The soft prompt composed of several learnable tokens receive comprehensive knowledge freely by generating appropriate semantics, while the hard prompt is a manually designed natural language sentence used to guide PLM in reasoning about news authenticity. Specifically, two tokens <s+>, <s-> with randomly initialized learnable embeddings $[\boldsymbol{s}^+, \boldsymbol{s}^-] \in \mathbb{R}^{2 \times d}$ are set as prefix soft prompt to receive the positively and negatively correlated knowledge, respectively. The soft prompt embedding after receiving knowledge is:

$$\mathbf{S} = \frac{1}{2}\left(\left[\boldsymbol{s}^+, \boldsymbol{s}^-\right] + \mathbf{E}^c\right) \quad (7)$$

While for postfix hard prompt, we test and adopt a cloze-style natural language sentence specialized for FND task, e.g., *"The veracity of the following news is [MASK]."*. The hard prompt embedding is denoted as $\mathbf{H} = \left[\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{[\text{MASK}]}, \ldots, \boldsymbol{h}_{n_h}\right] \in \mathbb{R}^{n_h \times d}$ where n_h is the number of hard prompt tokens, and $\boldsymbol{h}_{[\text{MASK}]}$ is the embedding of [MASK] token. It is got from the embedding layer of a tunable PLM called PLM-T.

The token embeddings of the given news $\mathcal{X}$, i.e., $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_x}] \in \mathbb{R}^{n_x \times d}$, is also got from the embedding layer of PLM-T, where n_x is the number of news tokens. The final hybrid prompt template $\mathbf{I}_{\text{prompt}}$ is:

$$\mathbf{I}_{\text{prompt}} = [\mathbf{S}, \mathbf{H}, \mathbf{X}] \quad (8)$$

4.2.2 Veracity Prediction

$\mathbf{I}_{\text{prompt}}$ is then sent to transformer layers of PLM-T to predict news veracity. Specifically, the output embedding $\boldsymbol{o}_{[\text{MASK}]} \in \mathbb{R}^d$ of [MASK] token is obtained as:

$$\boldsymbol{o}_{[\text{MASK}]} = \text{PLM-T}\left(\mathbf{I}_{\text{prompt}}\right) \quad (9)$$

Its vocabulary distribution $\boldsymbol{v}_{[\text{MASK}]} \in \mathbb{R}^{|\mathcal{V}|}$ is got by sending $\boldsymbol{o}_{[\text{MASK}]}$ to the head function of PLM, where $\mathcal{V}$ is the vocabulary of PLM. We manually define the vocabulary subsets $\mathcal{V}_* = \{\mathcal{V}_0, \mathcal{V}_1\}$, where $\mathcal{V}_0$ contains words about true, $\mathcal{V}_1$ contains words about fake. Then the probability of each label $y \in [0, 1]$ for a given news $\mathcal{X}$ is calculated as:

$$P(y \mid \mathcal{X}) = \frac{\exp\left(\boldsymbol{v}_{[\text{MASK}]}\left(\mathcal{V}_y\right)\right)}{\sum_{\mathcal{V}_i \in \mathcal{V}_*} \exp\left(\boldsymbol{v}_{[\text{MASK}]}\left(\mathcal{V}_i\right)\right)} \quad (10)$$

where $\boldsymbol{v}_{[\text{MASK}]}\left(\mathcal{V}_i\right)$ is sum of distribution scores of words in $\mathcal{V}_i$. Finally, the cross-entropy loss in each domain is as below, where $* \in \{s, t\}$ and $\mathcal{D}_*$ is the source or target domain dataset:

$$\mathcal{L}_{CE}^* = -\sum_{\left(\mathcal{X}_i^*, y_i^*\right) \in \mathcal{D}_*} y_i^* \log\left(P\left(y_i^* \mid \mathcal{X}_i^*\right)\right) \quad (11)$$

4.2.3 Adversarial Contrastive Training

In domain adaptive few-shot scenario, the detection performance is inherently determined by the quantity and quality of target domain samples, which are limited and may suffer from noises. To alleviate this problem, adversarial samples are generated by adding each target domain sample with a worst-case perturbation, i.e., a normed noisy vector towards the gradient direction that maximizes the loss $\mathcal{L}_{CE}^t$:

$$\boldsymbol{o}_{\text{[MASK],adv}}^t = \boldsymbol{o}_{\text{[MASK]}}^t + \frac{\nabla \mathcal{L}_{CE}^t}{\|\nabla \mathcal{L}_{CE}^t\|} \tag{12}$$

where $\nabla \mathcal{L}_{CE}^t$ is the first-order gradient of $\mathcal{L}_{CE}^t$, which is approximated by the Fast Gradient Value (Rozsa et al., 2016) method. The adversarial samples serve as additional target domain samples in training.

To facilitate PLM modeling the domain-invariant news-knowledge interaction pattern, we adopt contrastive loss function to reduce the inter-domain discrepancy by explicitly pulling close the output [MASK] embedding of news with the same label from target and source domains, respectively:

$$\mathcal{L}_{CL}^t = -\frac{1}{N \times M} \sum_{(\mathcal{X}_i^t, y_i^t) \in \mathcal{D}_t} \sum_{(\mathcal{X}_j^s, y_j^s) \in \mathcal{D}_s} \mathbb{1}_{[y_j^t = y_i^s]} \log \frac{\exp\left(S\left(\boldsymbol{o}_i^t, \boldsymbol{o}_j^s\right)/\tau\right)}{\sum_{(\mathcal{X}_k^s, y_k^s) \in \mathcal{D}_s} \exp\left(S\left(\boldsymbol{o}_i^t, \boldsymbol{o}_k^s\right)/\tau\right)} \tag{13}$$

where $S(\cdot)$ is cosine similarity, τ is a temperature parameter. Similarly, another contrastive loss is utilized to reduce the intra-class discrepancy for abundant source domain samples:

$$\mathcal{L}_{CL}^s = -\frac{1}{M \times (M-1)} \sum_{(\mathcal{X}_i^s, y_i^s) \in \mathcal{D}_s} \sum_{(\mathcal{X}_j^s, y_j^s) \in \mathcal{D}_s} \mathbb{1}_{[i \neq j]} \mathbb{1}_{[y_i^s = y_j^s]} \log \frac{\exp\left(S\left(\boldsymbol{o}_i^s, \boldsymbol{o}_j^s\right)/\tau\right)}{\sum_{(\mathcal{X}_k^s, y_k^s) \in \mathcal{D}_s} \mathbb{1}_{[i \neq k]} \exp\left(S\left(\boldsymbol{o}_i^s, \boldsymbol{o}_k^s\right)/\tau\right)} \tag{14}$$

The final loss of our model is then formulated as below, where α is a trade-off parameter:

$$\mathcal{L} = \sum_{* \in \{s,t\}} \alpha \mathcal{L}_{CE}^* + (1-\alpha) \mathcal{L}_{CL}^* \tag{15}$$

5 Experiments

5.1 Experiment Setup

Dataset. Three datasets are utilized to implement the experiments in domain adaptive few-shot setting. Snopes (Popat et al., 2017) is a domain-agnostic dataset providing news in various domains and is adopted as source domain dataset. Politifact (Shu et al., 2020) is a dataset specialized for US political system. CoAID (Cui and Lee, 2020) is a healthcare dataset containing COVID-19 related news. They are domain-specific datasets used as target domain datasets. The statistics of the datasets are reported in Table 2 in Appendix A.1.

Baseline. The COOL is compared with several groups of models suitable for domain adaptive few-shot FND, which include neural network-based models: **TextCNN** (Chen, 2015) and **Bi-LSTM** (Bahad et al., 2019); knowledge enhanced neural network-based model **KAN** (Dun et al., 2021); PLM-based models: **FT** (Liu et al., 2019), **ACLR** (Lin et al., 2022), **PET** (Schick and Schütze, 2021), **Soft-PT** (Li and Liang, 2021) and **RPL** (Lin et al., 2023); knowledge enhanced PLM-based model **KPL** (Jiang et al., 2022). The baseline methods are elaborated in Appendix A.2.

Implementation Details. We use Pytorch to implement our model [1]. For domain adaptive few-shot FND, the source domain dataset and a randomly selected K-shot subset of target domain dataset are available for model training, where $K \in \{2, 4, 8, 16\}$. The rest part of target domain dataset is used as test set to evaluate the detection performance. *Acc.* (Accuracy) and *F1* (Macro F1 score) are adopted for evaluating the performance, which have been widely used in previous works (Zhang et al., 2024; Lin et al., 2023). More details of implementation can be found in Appendix A.3.

5.2 Main Results

The comparison results are reported in Table 1. It is shown that our proposed COOL consistently achieves the best performance in all settings, with average improvements of 2.14% and 4.16% compared to the second-best method on Politifact and CoAID, respectively. Specifically, COOL performs better than all PLM-based methods, confirming the effectiveness of external knowledge in improving PLM in domain adaptive few-shot FND. The superiority of COOL over KPL, another knowledge

[1] The code will be released upon publication.

Table 1: Comparisons of different models on domain adaptive few-shot FND task. The best results are in boldface and the second-best results are underlined.

Target (Source)	# Shot	Metric	TextCNN	Bi-LSTM	KAN	FT	ACLR	PET	Soft-PT	RPL	KPL	COOL
Politifact (Snopes)	2	*Acc.*	0.5730	0.5730	0.5732	0.6171	<u>0.6388</u>	0.6208	0.5400	0.6316	0.6232	**0.6575**
		F1	0.5728	0.5710	0.5669	0.5947	<u>0.6375</u>	0.5982	0.5180	0.6196	0.6088	**0.6505**
	4	*Acc.*	0.6070	0.6093	0.5741	0.6350	0.6578	0.6660	0.5698	<u>0.6828</u>	0.6630	**0.7010**
		F1	0.5952	0.5965	0.5674	0.6210	0.6550	0.6312	0.5454	<u>0.6762</u>	0.6487	**0.6843**
	8	*Acc.*	0.6348	0.6490	0.6155	0.6691	<u>0.7849</u>	0.6793	0.6443	0.7742	0.7165	**0.7869**
		F1	0.6060	0.6202	0.6117	0.6683	<u>0.7243</u>	0.6754	0.6326	0.7072	0.7100	**0.7767**
	16	*Acc.*	0.6674	0.6963	0.6229	0.7556	0.7882	0.7775	0.6892	<u>0.8394</u>	0.8129	**0.8430**
		F1	0.6587	0.6932	0.6222	0.7381	0.7776	0.7764	0.6882	<u>0.8296</u>	0.8084	**0.8329**
CoAID (Snopes)	2	*Acc.*	0.3522	0.4129	0.4438	0.4626	0.5323	0.4315	0.4882	0.5433	<u>0.5494</u>	**0.6022**
		F1	0.2879	0.3284	0.3538	0.3615	0.4015	0.3616	0.3620	<u>0.4169</u>	0.3887	**0.4470**
	4	*Acc.*	0.4009	0.4915	0.4845	0.4927	0.5527	0.5960	0.5376	0.6916	<u>0.7261</u>	**0.7316**
		F1	0.3239	0.3765	0.3780	0.3930	0.4341	0.4494	0.3988	0.5086	<u>0.5334</u>	**0.5513**
	8	*Acc.*	0.4647	0.5363	0.5335	0.5447	0.5997	0.6607	0.5944	0.7227	<u>0.7392</u>	**0.7409**
		F1	0.3695	0.4230	0.4142	0.4221	0.4654	0.4818	0.4002	0.5307	<u>0.5534</u>	**0.5578**
	16	*Acc.*	0.4991	0.5470	0.5754	0.6474	0.6341	0.7136	0.6332	0.7336	<u>0.7562</u>	**0.7937**
		F1	0.3856	0.4371	0.4399	0.4697	0.4815	0.5328	0.4692	0.5542	<u>0.5609</u>	**0.5900**

enhanced prompt learning method, is possibly because: (1) our model incorporates more comprehensive knowledge by extracting both structured and unstructured knowledge that are positively or negatively correlated with news. (2) we design an adversarial contrastive enhanced hybrid prompt learning framework which incorporates comprehensive knowledge flexibly with appropriately learned semantics and guides PLM in modeling domain-invariant news-knowledge interaction patterns. The detailed comparisons between baselines are discussed in Appendix B.

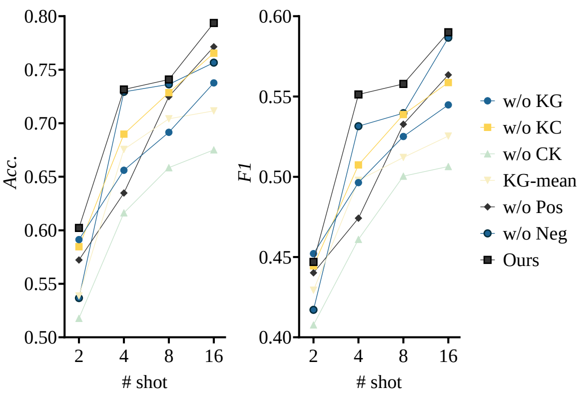


Figure 3: Ablation studies in comprehensive knowledge extraction module on CoAID, where “w/o KG” and “w/o KC” means removing structured knowledge retriever and unstructured knowledge retriever, respectively. “w/o CK” means removing the entire comprehensive knowledge extraction module. "KG-mean" means replacing the modulation mechanism with *mean pooling* in structured knowledge retriever. “w/o Pos” and “w/o Neg” means removing N-E$^+$ and N-E$^-$ attention, respectively.

5.3 Abaltion Study

Ablation studies are conducted to analyze the effects of the key designs in our model. The experimental results of ablation studies in comprehensive knowledge extraction module on CoAID are reported in Figure 3. If the structured knowledge retriever is removed, our model drops averagely 6.46% *Acc.* and 5.58% *F1*. When the unstructured knowledge retriever is removed, it drops averagely 4.71% *Acc.* and 4.30% *F1*. If the entire comprehensive knowledge extraction module is eliminated, our model reduces averagely by 12.85% *Acc.* and 12.40% *F1*. These results validate that both structured relational knowledge and unstructured descriptive knowledge are helpful for FND and they complement with each other to provide comprehensive knowledge. Moreover, if the modulation mechanism in structured knowledge retriever is replaced by *mean pooling*, COOL decreases averagely by 8.37% *Acc.* and 8.20% *F1*, which validates the effectiveness of modulation in attentively extracting structured knowledge and reducing noises. The efficacy of signed correlation-aware attention is further validated. Specifically, The model without N-E$^+$ attention drops averagely 6.91% *Acc.* and 6.12% *F1*, while the model without N-E$^-$ attention drops averagely 5.63% *Acc.* and 3.52% *F1*. This confirms that the knowledge either positively or negatively correlated with news provides critical evidences to verify news authenticity.

We also implement ablation experiments in hybrid prompt learning module on Politifact and report the results in Figure 4. When the prefix prompt is eliminated and the knowledge is directly concateand implementing contrastive training effectively overcomes the inter-domain discrepancy.

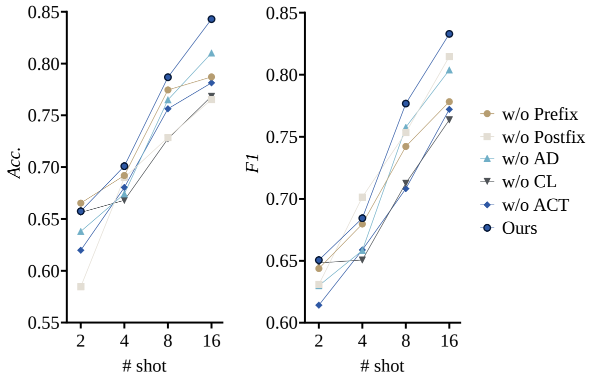


Figure 4: Ablation studies in hybrid prompt learning module on Politifact, where "w/o Prefix" and "w/o Postfix" means removing the prefix soft prompt and the postfix hard prompt, respectively. "w/o AD" and "w/o CL" means removing adversarial augmentation and contrastive training, respectively. "w/o ACT" means eliminates the adversarial contrastive training strategy.

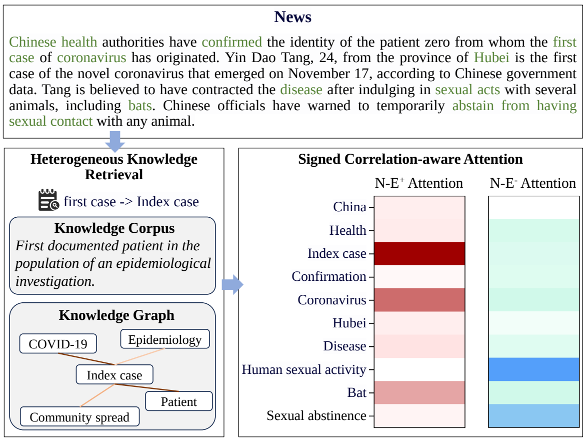


Figure 5: A real-world case from CoAID showing how COOL extracts comprehensive knowledge from external sources.

nated with postfix hard prompt, the model drops averagely 4.71% and 3.18% on *Acc.* and *F1*, which proves the advantage of incorporating knowledge with soft prompts to generate appropriate semantics. When the postfix hand-crafted prompt is removed, the model reduces averagely by 8.50% *Acc.* and 1.43% *F1*. This confirms the effect of task-specific hand-crafted prompt in guiding PLM in reasoning about news authenticity. The model removing adversarial augmentation drops averagely 3.36% *Acc.* and 3.22% *F1*, while the model removing contrastive training decreases averagely by 5.31% *Acc.* and 5.46% *F1*. If the entire adversarial contrastive training strategy is eliminated, the model reduces averagely by 4.95% *Acc.* and 6.36% *F1*. They demonstrate generating adversarial target domain samples improves model robustness and implementing contrastive training effectively overcomes the inter-domain discrepancy.

5.4 Case Study

To further explore how COOL extract comprehensive knowledge to enhance prompt learning in FND, we illustrate a case in CoAID in Figure 5. The news reports the index case of coronavirus in China. For every linked entity, we retrieve both structured and unstructured knowledge from heterogeneous external sources. For instance, "first case" mentioned in news is linked to entity "Index case", whose unstructured description is retrieved from knowledge corpus, and structured knowledge are filtered out from knowledge graph. The retrieved knowledge of all entities is then fed into the signed correlation-aware attention to extract knowledge that are positively or negatively correlated with news. Specifically, as shown in the heatmap, N-E$^+$ attention focus more on knowledge from entities that are highly positively correlated with news, such as "Coronavirus" and "Index case", while N-E$^-$ attention pay more attention to knowledge from entities that are not very (i.e., negatively) correlated with news, like "Human sexual activity". Intuitively, the extracted both positively and negatively correlated knowledge contribute to the authenticity judgement. The case study shows COOL can extract comprehensive and important knowledge for FND.

6 Conclusions

In this paper, we propose COOL, which extracts comprehensive knowledge from heterogeneous external sources and incorporates knowledge into hybrid prompt learning to verify news authenticity in domain adaptive few-shot scenario. The method is equipped with good expressiveness because: (i) we extract comprehensive knowledge that either positively or negatively correlate with news from both structured relational knowledge and unstructured description knowledge; (ii) we adopt hybrid prompt template which incorporates comprehensive knowledge freely by learned soft prompt and guides PLM in FND task by hand-crafted hard prompt; (iii) adversarial contrastive training is implemented to robustly model the domain-invariant news-knowledge interaction pattern. The extensive experiments on real-world datasets validate the effectiveness of COOL in domain adaptive few-shot FND and its capacity in incorporating comprehensive knowledge into prompt learning framework.

Limitations

Our work insists in injecting comprehensive knowledge into prompt learning for domain adaptive few-shot FND. However, in order to retrieve up-to-date external knowledge for FND in emerging domains, our model crawls Wikidata to obtain structured entity neighbors and unstructured entity descriptions, which can be time-consuming in pre-processing stage. Also, we find that the existing entity linking methods may overlook important news entities in some cases, which is a bottleneck for providing comprehensive knowledge information. Additionally, despite investigating prompt learning in domain adaptive FND, we do not discuss cross-domain adaptation with several source domains to improve detection performance in target domain.

References

Pritika Bahad, Preeti Saxena, and Raj Kamal. 2019. Fake news detection using bi-directional lstm-recurrent neural network. *Procedia Computer Science*, 165:74–82.

Shuanghao Bai, Min Zhang, Wanqi Zhou, Siteng Huang, Zhirong Luan, Donglin Wang, and Badong Chen. 2024. Prompt-based distribution alignment for unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 729–737.

Yahui Chen. 2015. Convolutional neural network for sentence classification. Master's thesis, University of Waterloo.

Limeng Cui and Dongwon Lee. 2020. Coaid: Covid-19 healthcare misinformation dataset. *arXiv preprint arXiv:2006.00885*.

Yaqian Dun, Kefei Tu, Chen Chen, Chunyan Hou, and Xiaojie Yuan. 2021. Kan: Knowledge-aware attention network for fake news detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 81–89.

Paolo Ferragina and Ugo Scaiella. 2010. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1625–1628.

Chunjiang Ge, Rui Huang, Mixue Xie, Zihang Lai, Shiji Song, Shuang Li, and Gao Huang. 2023. Domain adaptation via prompt learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Kunpeng Guo, Dennis Diefenbach, Antoine Gourru, and Christophe Gravier. 2023. Wikidata as a seed for web extraction. In *Proceedings of the ACM Web Conference 2023*, pages 2402–2411.

Lei Guo, Ziang Lu, Junliang Yu, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2024. Prompt-enhanced federated content representation learning for cross-domain recommendation. In *Proceedings of the ACM on Web Conference 2024*, WWW '24, page 3139–3149, New York, NY, USA. Association for Computing Machinery.

Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. 2021. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5149–5169.

Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 754–763.

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.

Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications*, 80(8):11765–11788.

Jiho Kim, Sungjin Park, Yeonsu Kwon, Yohan Jo, James Thorne, and Edward Choi. 2023. Factkg: Fact verification via reasoning on knowledge graphs. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16190–16206.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Jingqiu Li, Lanjun Wang, Jianlin He, Yongdong Zhang, and Anan Liu. 2023. Improving rumor detection by class-based adversarial domain adaptation. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, page 6634–6642, New York, NY, USA. Association for Computing Machinery.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4582–4597.

Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Chen Guang. 2022. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2543–2556.

Hongzhan Lin, Pengyao Yi, Jing Ma, Haiyun Jiang, Ziyang Luo, Shuming Shi, and Ruifang Liu. 2023. Zero-shot rumor detection with propagation structure via prompt learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5213–5221.

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. *arXiv preprint arXiv:1907.11692*.

Jing Ma, Chen Chen, Chunyan Hou, and Xiaojie Yuan. 2023. Kapalm: Knowledge graph enhanced language models for fake news detection. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3999–4009.

Nikhil Mehta, María Leonor Pacheco, and Dan Goldwasser. 2022. Tackling fake news detection by continually improving social context representations using graph neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1363–1380.

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.

Muhammad F Mridha, Ashfia Jannat Keya, Md Abdul Hamid, Muhammad Mostafa Monowar, and Md Saifur Rahman. 2021. A comprehensive review on fake news detection with deep learning. *IEEE access*, 9:156151–156170.

Qiong Nan, Danding Wang, Yongchun Zhu, Qiang Sheng, Yuhui Shi, Juan Cao, and Jintao Li. 2022. Improving fake news detection of influential domain via domain-and instance-level transfer. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2834–2848.

Yi Ouyang, Peng Wu, and Li Pan. 2022. Asymmetrical context-aware modulation for collaborative filtering recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1595–1604.

Shichao Pei, Qiannan Zhang, and Xiangliang Zhang. 2023. Few-shot low-resource knowledge graph completion with reinforced task generation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7252–7264.

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1003–1012.

Hongyan Ran and Caiyan Jia. 2023. Unsupervised cross-domain rumor detection with contrastive learning and cross-attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13510–13518.

Andras Rozsa, Ethan M Rudd, and Terrance E Boult. 2016. Adversarial diversity and hard positive generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 25–32.

Timo Schick and Hinrich Schütze. 2021. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 255–269.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706.

Mengzhu Sun, Xi Zhang, Jianqiang Ma, Sihong Xie, Yazheng Liu, and S Yu Philip. 2023. Inconsistent matters: A knowledge-guided dual-consistency network for multi-modal rumor detection. *IEEE Transactions on Knowledge and Data Engineering*.

Yu-Wun Tseng, Hui-Kuo Yang, Wei-Yao Wang, and Wen-Chih Peng. 2022. Kahan: knowledge-aware hierarchical attention network for fake news detection on social media. In *Companion Proceedings of the Web Conference 2022*, pages 868–875.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Jiaying Wu, Shen Li, Ailin Deng, Miao Xiong, and Bryan Hooi. 2023. Prompt-and-align: Prompt-based social alignment for few-shot fake news detection. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, page 2726–2736, New York, NY, USA. Association for Computing Machinery.

Liang Xiao, Qi Zhang, Chongyang Shi, Shoujin Wang, Usman Naseem, and Liang Hu. 2024. Msynfd: Multi-hop syntax aware fake news detection. In *Proceedings of the ACM on Web Conference 2024*, WWW '24, page 4128–4137, New York, NY, USA. Association for Computing Machinery.

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.

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 (Volume 1: Long Papers)*, pages 5223–5239.

Litian Zhang, Xiaoming Zhang, Ziyi Zhou, Feiran Huang, and Chaozhuo Li. 2024. Reinforced adaptive knowledge learning for multimodal fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16777–16785.

Lulu Zhao, Fujia Zheng, Weihao Zeng, Keqing He, Ruotong Geng, Huixing Jiang, Wei Wu, and Weiran Xu. 2022. Adpl: Adversarial prompt-based domain adaptation for dialogue summarization with knowledge disentanglement. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 245–255.

Anni Zou, Zhuosheng Zhang, and Hai Zhao. 2023. Decker: Double check with heterogeneous knowledge for commonsense fact verification. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 11891–11904.

A Experiment Setup Details

A.1 Datasets Statistics

Three datasets are used to conduct domain-adaptive few-shot FND experiments, where Snopes (Popat et al., 2017) is a domain-agnostic dataset which is extracted from a fact-checking website [2] providing various news articles and corresponding labels. Politifact (Shu et al., 2020) is a political related dataset collected from another fact-checking website [3] specialized for US political system. CoAID (Cui and Lee, 2020) is a healthcare dataset containing COVID-19 related news on websites and social platforms. In our experiments, the domain-agnostic dataset Snopes is adopted as source domain dataset while the domain-specific datasets, Politifact and CoAID, are used as target domain datasets. We filter out the articles whose URL is no longer accessible. The statistical details of the datasets after prepossessing are summarized in Table 2.

[2]https://www.snopes.com/

[3]https://www.politifact.com/

Table 2: Statistics of the datasets

Datasets	Snopes	Politifact	CoAID
# News	710	886	2807
# Real	430	517	2652
# Fake	280	369	155
Avg. # words	690	1361	78
Avg. # entities	126	239	18

A.2 Baseline Methods

To evaluate the performance of our proposed model, we compare COOL with several groups of models to conduct domain adaptive few-shot FND experiments:

The first group of models is neural network-based models:

TextCNN (Chen, 2015): This method uses convolutional neural networks with multiple filter widths to extract text features which are further fed into pooling layer and fully connected layer for classification.

Bi-LSTM (Bahad et al., 2019): This method utilizes bi-directional long short-term memory which exploits text sequence from front-to-back and back-to-front and recurrent neural network for FND.

The second group of models is knowledge enhanced neural network-based models:

KAN (Dun et al., 2021): This method proposes a knowledge-aware attention network for FND by extracting external knowledge from knowledge graph that are most relevant to news semantics with attention mechanism.

The third group of models is PLM-based models:

FT (Liu et al., 2019): This is the standard fine-tuning method built on top of RoBERTa by feeding [CLS] embedding into task-specific linear layers to predict news veracity.

ACLR (Lin et al., 2022): This is a state-of-the-art domain adaptive FND method that adapts features learned from rich source domain to low-resource target domain by developing adversarial augmentation mechanism and supervised contrastive training paradigm.

PET (Schick and Schütze, 2021): This is a prompt learning method that provides task-related hand-crafted prompt to reformulate input as cloze-style phrases to help PLM understand the given task.

Soft-PT (Li and Liang, 2021): This method uses learnable tokens to optimize a sequence of continuous task-specific vectors for prompt tuning instead of discrete prompt which is constrained to real words embeddings.

RPL (Lin et al., 2023): This is a state-of-the-art prompt learning-based method that generates adversarial augmentation examples and introduces a prototypical verbalizer paradigm with designed contrastive learning framework for detection task.

The fourth group of models is knowledge enhanced PLM-based models:

KPL (Jiang et al., 2022): This is a strong baseline, which applies prompt learning to FND and incorporates knowledge features extracted from entity sequence into learnable prompts.

A.3 Implementation Details

We further detail our implementation as follows: The mini-batch Adaptive Moment Estimation (Adam) (Kingma and Ba, 2014) is adopted as the optimizer, which can adaptively adjust the learning rate during the training phase. We utilize Tagme (Ferragina and Scaiella, 2010) as the entity linking method, while Wikidata (Guo et al., 2023) is used as external knowledge sources to crawl entity neighbors and entity descriptions. The hyper-parameter settings are as follows: training batch size is 16, hidden dimension of PLM is 768, the learning rate is 2e-5, the trade-off parameter α is 0.5, the temperature parameter τ is 0.1. For all baselines, the optimal hyper-parameter settings are determined either by our experiments or suggested by previous works to ensure the best performance. For fair comparisons, the base version of RoBERTa and Wikidata are used as PLM and external source for all needed methods, respectively, and the self-training and PLM ensemble for PET are not implemented, following previous work (Wu et al., 2023). As we address the situations where no propagation structure can be obtained, and the adopted source dataset Snopes doesn't have the propagation structure, the propagation position modeling and the response ranking for RPL are not implemented. We add the cross-entropy loss into the training of RPL which is not included in the original paper, confirming the model stability in our experiments. All of our experiments are run on one single NVIDIA RTX A6000 GPU. The reported comparative results are averaged from ten implementations with randomly choiced seeds.

B Comparative Analysis

Apart from the observation that COOL consistently outperforms baseline methods, there are more conclusions can be drawn from the comparative results shown in Table 1.

First, PLM-based methods generally outperform neural network-based methods on all experimental settings, which indicates the strong capacity of PLM in extracting semantic features of news to model fake news pattern. Among PLM-based methods, PET outperforms FT in most settings, which demonstrates the superiority of prompt learning over fine-tuning in domain adaptive scenario. However, Soft-PT generally performs worst in PLM-based methods, which indicates relying solely on randomly initialized soft prompts cannot effectively guide PLM in reasoning about task-related news authenticity.

Second, there are some findings in the comparison of knowledge enhanced models. As we can see in Table 1, different from many FND scenarios, knowledge enhanced neural network-based method performs generally worse than neural network-based methods in our experiments. This suggests that without non-trivial designs to overcome the inter-domain discrepancy, the news-knowledge interaction captured by knowledge enhanced neural network-based model is domain-specific which degenerates the model performance in emerging news domain. While for PLM-based models, KPL enhanced by knowledge outperforms PET suggests PLM's potential in incorporating knowledge to boost detection performance.

Third, KPL generally performs the second-best results in the experiments on Covid. This suggests that introducing knowledge information is especially helpful for FND in emerging news domains since CoAID may be more related to recent real-world knowledge that are not contained in PLM pre-training corpus. Moreover, ACLR and RPL generally perform the second-best results in the experiments on Politifact, showing that the adversarial augmentation and contrastive training equipe them with strong domain adaptive learning capacity.