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

Yi Ouyang[1,2], Peng Wu[1,2(✉)], and Li Pan[1,2]

[1] Shanghai Jiao Tong University, Shanghai, China
[2] Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai Jiao Tong University, Shanghai, China
{halaoyy,catking,panli}@sjtu.edu.cn

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in verifying emerging news, as emerging news often involves timely knowledge that may not be contained in the PLM's outdated prior knowledge.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.

Keywords: Fake news detection · Prompt learning · External knowledge · Few-shot Learning.

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 [18, 11, 19]. 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 downstream learning the same as

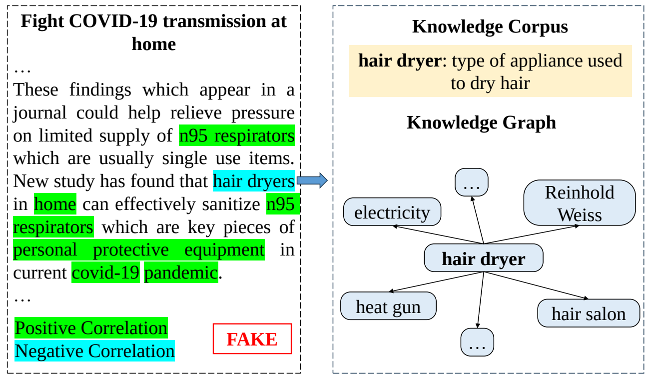


Fig. 1: A piece of real-world fake news about the prevention of Covid-19.

the pre-training process, achieves success in few-shot scenarios and has been used in various domain adaptive tasks [2, 7].

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 [27], to assist in domain adaptive FND. Most previous studies generally extract entity knowledge positively correlated with news [5, 30, 17]. However, the negatively correlated knowledge, i.e., entity knowledge not very correlated with news semantics, may also contribute 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 [17, 5]. 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 [10]. 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 extracting more comprehensive knowledge and capturing the domain-invariant news-knowledge interaction patterns.

To this end, we propose COOL for domain adaptive few-shot FND, which extracts comprehensive knowledge that is positively or negatively correlated with news from heterogeneous sources, and injects it into prompt learning by an adversarial contrastive enhanced hybrid prompt learning framework. Specifically, a comprehensive knowledge extraction module is proposed to first retrieve both structured and unstructured knowledge from external sources, and then filter both positively and negatively correlated knowledge via a signed correlation-aware attention. When retrieve news–related structured knowledge, a modulation mechanism is adopted instead of attention mechanism to avoid computational intractability. The filtered comprehensive knowledge is incorporated by a hybrid prompt learning framework, where prefix soft prompt composed of several learnable tokens receives knowledge representations flexibly, while postfix 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 datasets [18, 11, 19]. 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 [32, 14, 23], such as meta-learning improving domain adaptation by adjusting model parameters step by step across tasks [33, 20, 9], and contrastive learning reducing the inter-domain discrepancy by appropriate contrastive loss [32, 14, 23]. 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 [15]. However, 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 [5, 30, 17, 34]. Most existing methods leverage structured knowledge graph, e.g. ConceptNet [27] and YAGO [28], to capture the relational knowledge among entities for FND [17, 29]. Some studies benefit from abundant heterogeneous knowledge by exploiting both structured knowledge graph and unstructured knowledge corpus, e.g., Wikipedia corpus [35]. 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 [10] devises a knowledgeable prompt learning framework which incorporates the sequential knowledge entities into prompt template to predict the news veracity. Different from [10], 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 $\mathcal{D}_s = \{(\mathcal{X}_1^s, y_1^s), (\mathcal{X}_2^s, y_2^s), \ldots, (x_M^s, y_M^s)\}$ denote the source domain dataset 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 target domain dataset, 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

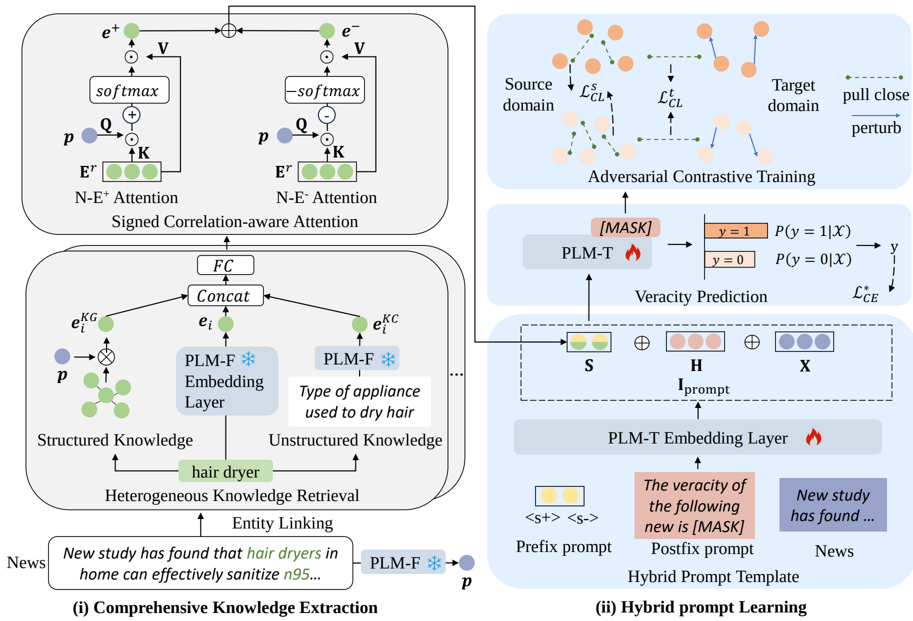


Fig. 2: Architecture of the proposed COOL model.

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

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 [6]. The identified entities are embedded by the embedding layer of a parameter-frozen PLM, i.e., $\mathbf{E} = \left[\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{|\mathcal{E}|}\right] \in \mathbb{R}^{|\mathcal{E}| \times d}$, 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 vectorize various information and 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". 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 [31]. 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[5, 30]. Instead, we adopt a modulation mechanism based on *product & max pooling* [21] 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:

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

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 [5, 29]. 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:

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

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:

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

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:

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

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:

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

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

The final comprehensive knowledge is set as $\mathbf{E}^{c} = [\boldsymbol{e}^{+}, \boldsymbol{e}^{-}] \in \mathbb{R}^{2\times d}$ for FND.

4.2 Hybrid Prompt Learning

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 sentence that is manually designed 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:

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

While for postfix hard prompt, we 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 $\mathbf{H} = \left[\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{[\mathrm{MASK}]}, \ldots, \boldsymbol{h}_{n_h}\right] \in \mathbb{R}^{n_h \times d}$ where n_h is the number of hard prompt tokens, and $\boldsymbol{h}_{[\mathrm{MASK}]}$ 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:

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

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

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

Its vocabulary distribution $\boldsymbol{v}_{[\mathrm{MASK}]} \in \mathbb{R}^{|\mathcal{V}|}$ is got by sending $\boldsymbol{o}_{[\mathrm{MASK}]}$ 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:

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

where $\boldsymbol{v}_{[\mathrm{MASK}]}\left(\mathcal{V}_i\right)$ 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:

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

Adversarial Contrastive Training. The detection performance of domain adaptive few-shot FND 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 $\mathcal{L}_{CE}^t$:

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

where $\nabla \mathcal{L}_{CE}^t$ is the first-order gradient of $\mathcal{L}_{CE}^t$, which is approximated by the Fast Gradient Value [24] 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:

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

The final loss 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. In real scenarios, FND models are usually trained on some domain-agnostic datasets composed of news from various domains, and are required to be applied in datasets of specific emerging domain. Thus, we adopt a domain-agnostic dataset Snopes [22] as the source domain dataset, and adopt two domain-specific datasets, i.e., Politifact [26] specialized for US political system, and CoAID [4] containing COVID-19 related news, as target domain datasets. The statistics of the datasets are reported in Table 1.

Table 1: 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

Baseline. The COOL is compared with several groups of models, which include neural network-based models: **TextCNN** [3] and **Bi-LSTM** [1]; knowledge enhanced neural network-based model **KAN** [5]; PLM-based models: **FT** [16], **ACLR** [14], **PET** [25], **Soft-PT** [13] and **RPL** [15]; knowledge enhanced PLM-based model **KPL** [10].

Implementation Details. We use Pytorch to implement our model [3]. 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 [34, 15]. The mini-batch Adaptive Moment Estimation (Adam) [12] is adopted as the optimizer, which can adaptively adjust the learning rate during the training phase. We utilize Tagme [6] as the entity linking method, while Wikidata [8] 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.

Table 2: 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.*	0.573	0.573	0.573	0.617	<u>0.638</u>	0.620	0.540	0.631	0.623	**0.657**
		F1	0.572	0.571	0.566	0.594	<u>0.637</u>	0.598	0.518	0.619	0.608	**0.650**
	4	*Acc.*	0.607	0.609	0.574	0.635	0.657	0.666	0.569	<u>0.682</u>	0.663	**0.701**
		F1	0.595	0.596	0.567	0.621	0.655	0.631	0.545	<u>0.676</u>	0.648	**0.684**
	8	*Acc.*	0.634	0.649	0.615	0.669	<u>0.784</u>	0.679	0.644	0.774	0.716	**0.786**
		F1	0.606	0.620	0.611	0.6683	<u>0.724</u>	0.675	0.632	0.707	0.710	**0.776**
	16	*Acc.*	0.667	0.696	0.622	0.755	0.788	0.777	0.689	<u>0.839</u>	0.812	**0.843**
		F1	0.658	0.693	0.622	0.738	0.777	0.776	0.688	<u>0.829</u>	0.808	**0.832**
CoAID (Snopes)	2	*Acc.*	0.352	0.412	0.443	0.462	0.532	0.431	0.488	0.543	<u>0.549</u>	**0.602**
		F1	0.287	0.328	0.353	0.361	0.401	0.361	0.362	<u>0.416</u>	0.388	**0.447**
	4	*Acc.*	0.400	0.491	0.484	0.492	0.552	0.596	0.537	0.691	<u>0.726</u>	**0.731**
		F1	0.323	0.376	0.378	0.393	0.434	0.449	0.398	0.508	<u>0.533</u>	**0.551**
	8	*Acc.*	0.464	0.536	0.533	0.544	0.599	0.660	0.594	0.722	<u>0.739</u>	**0.740**
		F1	0.369	0.423	0.414	0.422	0.465	0.481	0.400	0.530	<u>0.553</u>	**0.557**
	16	*Acc.*	0.499	0.547	0.575	0.647	0.634	0.713	0.633	0.733	<u>0.756</u>	**0.793**
		F1	0.385	0.437	0.439	0.469	0.481	0.532	0.469	0.554	<u>0.560</u>	**0.590**

[3] https://anonymous.4open.science/r/COOL-anonymous

5.2 Main Results

The comparison results are reported in Table 2. 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 enhanced prompt learning method, is possibly because: (1) our model incorporates more comprehensive knowledge by extracting both structured and unstructured knowledge that positively or negatively correlates with news. (2) we design an adversarial contrastive enhanced hybrid prompt learning framework which incorporates comprehensive knowledge flexibly with learned semantics and guides PLM in modeling domain-invariant news-knowledge interaction patterns.

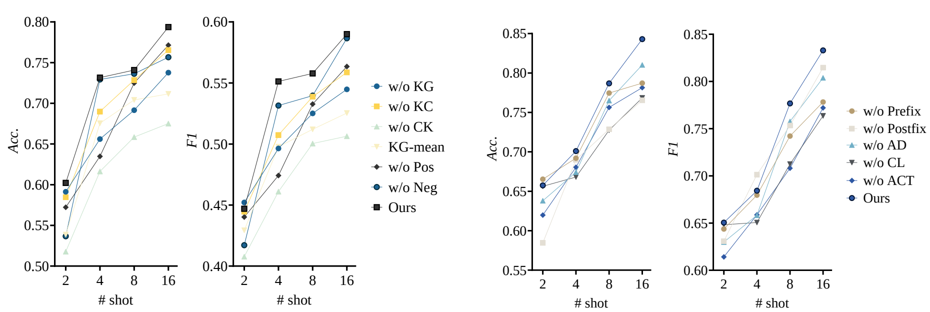
The experimental setup where the model can be trained on widely collected domain-agnostic datasets and then applied on emerging domain-specific datasets simulates most real-world detection scenarios. Additionally, to validate the model generalizability, COOL is also compared with the two competitive baselines, i.e., RPL and KPL, in scenarios where domain-specific dataset Politifact and CoAID are set as the source and target datasets for each other, and the results are reported in Table 3. Specifically, when CoAID is set as the source domain dataset and Politifact is set as the target domain dataset, COOL averagely outperforms the second-best with 6.35% *Acc.* and 6.91% *F1*. When Politifact and CoAID are set as source and domain datasets respectively, COOL averagely outperforms the second-best with 5.42% *Acc.* and 6.67% *F1*. These experimental results further validates the model generalizability in domain adaptive few-shot setting.

Table 3: Comparisons of COOL and baselines where source and target domain datasets are both domain specific datasets. The best results are in boldface and the second-best results are underlined.

Target (Source)	#Shot	Metric	RPL	KPL	COOL
Politifact (CoAID)	2	*Acc.*	<u>0.632</u>	0.579	**0.681**
		F1	<u>0.626</u>	0.571	**0.679**
	4	*Acc.*	<u>0.708</u>	0.666	**0.767**
		F1	<u>0.705</u>	0.653	**0.762**
	8	*Acc.*	0.745	<u>0.785</u>	**0.807**
		F1	0.737	<u>0.758</u>	**0.790**
CoAID (Politifact)	2	*Acc.*	0.5298	<u>0.540</u>	**0.588**
		F1	0.423	<u>0.428</u>	**0.489**
	4	*Acc.*	0.649	<u>0.695</u>	**0.701**
		F1	0.489	<u>0.499</u>	**0.521**
	8	*Acc.*	0.699	<u>0.713</u>	**0.759**
		F1	0.521	<u>0.540</u>	**0.548**

5.3 Abaltion Study

Ablation studies are conducted to analyze the effects of the key designs in COOL. The ablation studies in comprehensive knowledge extraction module on CoAID are reported in Figure 3a. 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, COOL 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 without N-E$^-$ attention drops averagely 5.63% *Acc.* and 3.52% *F1*. This confirms either positively or negatively correlated knowledge provides critical evidences to verify news authenticity.



(a) Ablation studies in comprehensive knowledge extraction module on CoAID, where "w/o KG" and "w/o KC" means removing structured 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.

(b) Ablation studies in hybrid prompt learning module on Politifact, where "w/o Prefix" and "w/o Postfix" stand for removing the prefix soft prompt and the postfix hard prompt, respectively. "w/o AD" means removing adversarial augmentation, while "w/o CL" means removing contrastive training. "w/o ACT" means eliminating the adversarial contrastive training strategy.

Fig. 3: Ablation studies in comprehensive knowledge extraction module and hybrid prompt learning module.

Ablation experiments in hybrid prompt learning module on Politifact and are reported in Figure 3b. When the prefix prompt is eliminated and the knowledge is directly concatenated 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.

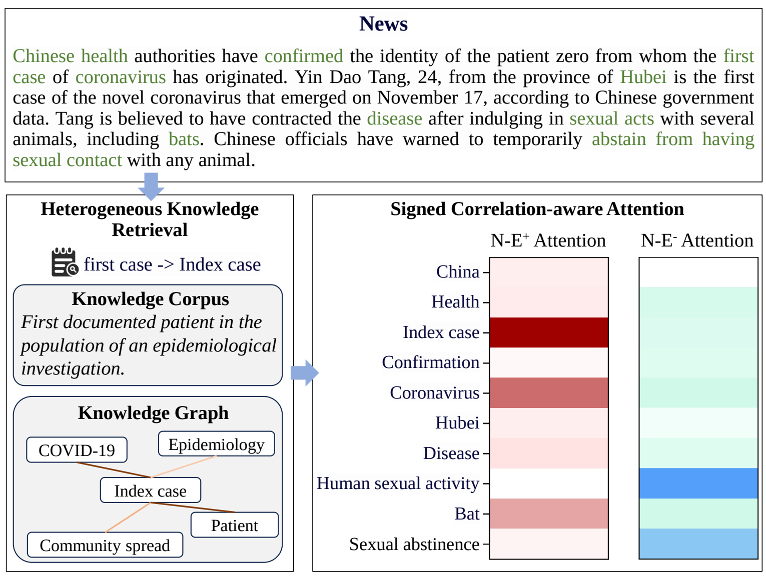


Fig. 4: A real-world case from CoAID showing how COOL extracts comprehensive knowledge.

5.4 Case Study

To further explore how COOL extract comprehensive knowledge for FND, we illustrate a case in CoAID in Figure 4. The news reports the index case of coronavirus. For every linked entity, we retrieve both structured and unstructured knowledge from heterogeneous 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 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 “sexual abstinence”. Intuitively, the extracted both positively and negatively correlated knowledge contribute to the authenticity judgment. The case shows COOL can extract important comprehensive 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 descriptive 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. Extensive experiments validate the effectiveness of COOL for FND and its capacity in incorporating comprehensive knowledge into prompt learning framework.

Acknowledgments

This work is supported by the National Key Research and Development Plan in China (No. 2023YFC3306100), the National Natural Science Foundation of China (No. 62172278, 62002219).

References

1. Bahad, P., Saxena, P., Kamal, R.: Fake news detection using bi-directional lstm-recurrent neural network. Procedia Computer Science **165**, 74–82 (2019)
2. Bai, S., Zhang, M., Zhou, W., Huang, S., Luan, Z., Wang, D., Chen, B.: Prompt-based distribution alignment for unsupervised domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 729–737 (2024)
3. Chen, Y.: Convolutional neural network for sentence classification. Master’s thesis, University of Waterloo (2015)
4. Cui, L., Lee, D.: Coaid: Covid-19 healthcare misinformation dataset. arXiv preprint arXiv:2006.00885 (2020)

5. Dun, Y., Tu, K., Chen, C., Hou, C., Yuan, X.: Kan: Knowledge-aware attention network for fake news detection. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 81–89 (2021)
6. Ferragina, P., Scaiella, U.: 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. pp. 1625–1628 (2010)
7. Ge, C., Huang, R., Xie, M., Lai, Z., Song, S., Li, S., Huang, G.: Domain adaptation via prompt learning. IEEE Transactions on Neural Networks and Learning Systems (2023)
8. Guo, K., Diefenbach, D., Gourru, A., Gravier, C.: Wikidata as a seed for web extraction. In: Proceedings of the ACM Web Conference 2023. pp. 2402–2411 (2023)
9. Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A.: Meta-learning in neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(9), 5149–5169 (2021)
10. Jiang, G., Liu, S., Zhao, Y., Sun, Y., Zhang, M.: Fake news detection via knowledgeable prompt learning. Information Processing & Management **59**(5), 103029 (2022)
11. Kaliyar, R.K., Goswami, A., Narang, P.: Fakebert: Fake news detection in social media with a bert-based deep learning approach. Multimedia tools and applications **80**(8), 11765–11788 (2021)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Li, X.L., Liang, P.: 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). pp. 4582–4597 (2021)
14. Lin, H., Ma, J., Chen, L., Yang, Z., Cheng, M., Guang, C.: Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. In: Findings of the Association for Computational Linguistics: NAACL 2022. pp. 2543–2556 (2022)
15. Lin, H., Yi, P., Ma, J., Jiang, H., Luo, Z., Shi, S., Liu, R.: Zero-shot rumor detection with propagation structure via prompt learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 5213–5221 (2023)
16. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
17. Ma, J., Chen, C., Hou, C., Yuan, X.: Kapalm: Knowledge graph enhanced language models for fake news detection. In: Findings of the Association for Computational Linguistics: EMNLP 2023. pp. 3999–4009 (2023)
18. Mehta, N., Pacheco, M.L., Goldwasser, D.: 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). pp. 1363–1380 (2022)
19. Mridha, M.F., Keya, A.J., Hamid, M.A., Monowar, M.M., Rahman, M.S.: A comprehensive review on fake news detection with deep learning. IEEE access **9**, 156151–156170 (2021)
20. Nan, Q., Wang, D., Zhu, Y., Sheng, Q., Shi, Y., Cao, J., Li, J.: Improving fake news detection of influential domain via domain-and instance-level transfer. In: Proceedings of the 29th International Conference on Computational Linguistics. pp. 2834–2848 (2022)

21. Ouyang, Y., Wu, P., Pan, L.: Asymmetrical context-aware modulation for collaborative filtering recommendation. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 1595–1604 (2022)
22. Popat, K., Mukherjee, S., Strötgen, J., Weikum, G.: 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. pp. 1003–1012 (2017)
23. Ran, H., Jia, C.: Unsupervised cross-domain rumor detection with contrastive learning and cross-attention. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 13510–13518 (2023)
24. Rozsa, A., Rudd, E.M., Boult, T.E.: Adversarial diversity and hard positive generation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 25–32 (2016)
25. Schick, T., Schütze, H.: 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. pp. 255–269 (2021)
26. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: 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 (2020)
27. Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of the AAAI conference on artificial intelligence. vol. 31 (2017)
28. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Proceedings of the 16th international conference on World Wide Web. pp. 697–706 (2007)
29. Sun, M., Zhang, X., Ma, J., Xie, S., Liu, Y., Philip, S.Y.: Inconsistent matters: A knowledge-guided dual-consistency network for multi-modal rumor detection. IEEE Transactions on Knowledge and Data Engineering (2023)
30. Tseng, Y.W., Yang, H.K., Wang, W.Y., Peng, W.C.: Kahan: knowledge-aware hierarchical attention network for fake news detection on social media. In: Companion Proceedings of the Web Conference 2022. pp. 868–875 (2022)
31. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
32. Yue, Z., Zeng, H., Kou, Z., Shang, L., Wang, D.: 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. pp. 2423–2433 (2022)
33. Yue, Z., Zeng, H., Zhang, Y., Shang, L., Wang, D.: 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). pp. 5223–5239 (2023)
34. Zhang, L., Zhang, X., Zhou, Z., Huang, F., Li, C.: Reinforced adaptive knowledge learning for multimodal fake news detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 16777–16785 (2024)
35. Zou, A., Zhang, Z., Zhao, H.: Decker: Double check with heterogeneous knowledge for commonsense fact verification. In: Findings of the Association for Computational Linguistics: ACL 2023. pp. 11891–11904 (2023)