Computer Science and Information Systems 22(1):33–57 https://doi.org/10.2298/CSIS240320067Z

Topic-oriented Sarcasm Detection via Entity Knowledge-based Prompt Learning

Yuhao Zhou[1], Shunxiang Zhang[1,*], Caiqin Wang[1], Yanhui Wang[1], Xiaolong Wang[1] and KuanChing Li[2,*]

[1] School of Computer Science and Engineering,
Anhui University of Science & Technology, 232001 Huainan, China
1303496845@qq.com
sxzhang@aust.edu.cn
1976499113@qq.com
2457172801@qq.com
1556282598@qq.com
[2] Department of Computer Science and Information Engineering (CSIE),
Providence University, 43301 Taizhong, Taiwan
kuancli@pu.edu.tw

Abstract. The extensive spread of sarcasm on social media has attracted great attention to sarcasm detection. Topic-oriented sarcasm detection aims to determine the sarcastic tendency of a comment on a specific topic. Existing methods focus on using topics as contextual information to enhance comprehension of comment semantics. However, when topics and comments contain entities with knowledge information, accurately understanding the comment semantics becomes challenging. To this end, we investigate an Entity Knowledge-based Prompt Learning (EKPL) model that combines prompt learning and entity knowledge from knowledge graphs for topic-oriented sarcasm detection. Specifically, we use prompt learning to transform topic-oriented sarcasm detection from a classification task to a mask prediction task, while we incorporate entity knowledge into the prompt representation to enhance the expressiveness of its predictive mask words and the model's understanding of text semantics. Experimental results on the public *ToSarcasm*[†] dataset illustrate that our EKPL model has a significant performance in topic-oriented sarcasm detection task.

Keywords: Sarcasm detection, Entity knowledge, Prompt Learning, Knowledge graph, EKPL.

1. Introduction

As a crucial medium for online communication, social media platforms' real-time nature and convenience allow people to stay updated on trending topics and make comments at

---

[*] Corresponding authors

[†] https://github.com/HITSZ-HLT/ToSarcasm/tree/main

any time. This has led to rich and diverse expressions of social media comments, such as irony, sarcasm, and humor. Sarcasm is a rhetorical device widely used by people and is characterized by indirectness [1]. The characteristic often leads to a significant contrast between the literal meaning of comments and the true emotions of users, resulting in misunderstandings of the comments. This brings huge challenges to user opinion mining and sentiment analysis [2].

With the advancement of deep learning, numerous deep learning-based methods have been proposed [3-4], greatly improving the detection performance. Most of the current research on sarcasm detection for social media comments focuses on sentence-level text sarcasm detection. However, social media users' comments typically revolve around specific topics or events. A topic may have multiple comments, which are intended to describe or subjectively reflect users' opinions on the topic. In light of this phenomenon, the study [5] proposed a topic-oriented sarcasm detection task. This task requires judging whether a comment is a sarcastic expression (i.e., sarcasm or non-sarcasm) on a specific topic, which is different from traditional sentence-level text sarcasm detection. To solve this task, they built a topic-oriented sarcastic expression prompt learning model. This model leverages prompt learning to utilize the Pretrained Language Model (PLM) and achieves better performance than models that utilize PLM based on features and fine-tuning, thereby effectively modeling the topic-oriented sarcasm detection task.

topic:"因不接种疫苗致麻疹爆发 纽约市郊进入紧急状态"
comment:"妥妥的，智商税"

topic:"汤姆·霍兰德发问全球汽车巨头们会成为另一个诺基亚吗？ "
comment:"汽油车就是切用切珍惜吧，二氧化碳排完了就没得排了，什么生物柴油乙醇汽油都是扯淡的，燃油车以后就是军用车特种车还有点价值，毕竟还有点能量密度高、后勤保障简单的优点"

Fig. 1. Two examples of topic-comment text pairs on *ToSarcam*

As a new paradigm for utilizing PLM, prompt learning aims to effectively use pre-training information by aligning the learning process with pre-training objectives. This approach overcomes the data starvation problem of fine-tuning. Previous PLM utilization methods, especially fine-tuning, have achieved great success under data-sufficient conditions, but they tend to perform poorly in low-resource scenarios. Unlike methods that utilize PLM to directly output class distributions based on features and fine-tuning, prompt learning methods mask specific label words and make predictions like a cloze problem. (e.g., *“这是一条针对话题1的<_>评论。”* This is a <_> comment for Topic 1.). This greatly reduces the disparity between pre-training and target task. Additionally, in low-resource environments with a limited number of training examples, prompt learning has also shown significant effectiveness [6-8]. Therefore, prompt learning is anticipated to enhance topic-oriented sarcasm detection task in real-world scenarios.

In addition, knowledge graphs can benefit topic-oriented sarcasm detection task through entity representation. Generally, topics are condensed, comments are subjectively tendentious, and both are filled with entities. These entities may appear in a variety of forms, such as aliases, abbreviations, and alternative spellings. Figure 1 shows two topic-comment text pairs in the *ToSarcam* dataset, which include entities such as *"纽约市郊"* (suburbs of New York City), *"智商税"* (IQ tax), *"汤姆•霍兰德"* (Tom Hollander), *"诺基亚"* (Nokia), and *"汽油车"* (gasoline car). In the comment *"妥妥的，智商税"* (Sure enough, IQ tax), the entity *"智商税"* (IQ tax) is a slang or metaphorical term. It refers to the consequences of not getting *"疫苗"* (vaccinated) leading to an outbreak of *"麻疹"* (measles) in the *"纽约市郊"* (suburbs of New York City). In another topic-comment text pair, the entity *"汽油车"* (gasoline car) is another name for the entity *"燃油车"* (fuel car), and they both refer to cars that need to use oil. In addition, the entity *"生物柴油"* (biodiesel) and the entity *"乙醇汽油"* (ethanol gasoline) refer to one type of diesel and gasoline respectively. These entities usually contain a lot of semantic information and background knowledge. Therefore, we can help the model understand these different entities by introducing external knowledge, thereby improving the model's accuracy in judging sarcastic comments.

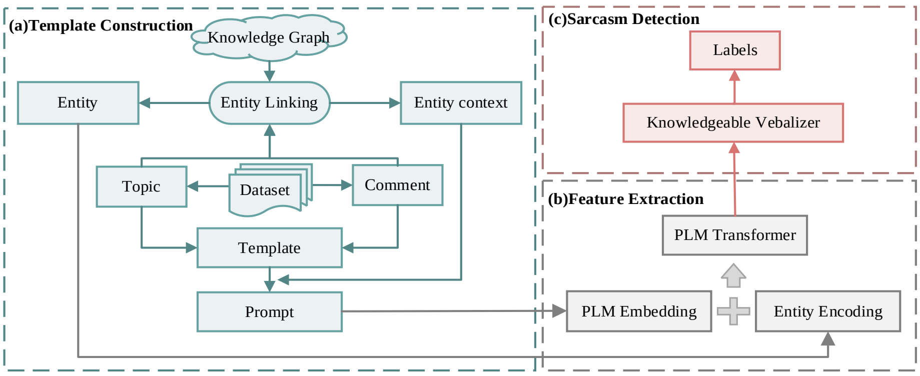

Fig. 2. The frame diagram of the EKPL model

For the model to understand these entities, we use the knowledge graph to obtain the corresponding entity information. The knowledge graph is a multi-relationship graph, which is composed of entity nodes and entity relationship edges. The relationship edges can describe the directional relationship between two entity nodes. The knowledge-level judgments and connections are beneficial to understanding the content of topics and comments because: (1) ambiguity can be avoided by linking each entity in topics and comments to its corresponding entity in the knowledge graph, and the problem of ambiguous entity mentions can be solved. (2) the knowledge graph can also provide more background information about entities, that is, entity context information. It helps to learn knowledge-level relationships among entities in topics and comments and improves the performance of sarcasm detection. Therefore, the incorporation of external knowledge is crucial for topic-oriented sarcasm detection.

In this paper, we propose an entity knowledge-based prompt learning model for the topic-oriented sarcasm detection task. The frame diagram of the model is shown in Figure 2. First, we design a topic-oriented prompt template for topics. Then we identify entities in topics and comments and obtain the corresponding entities using the knowledge graph *Wikidata* [9]. Next, we extract each entity's directly connected neighbor entities in the knowledge graph as entity context information. Finally, the entity context information and entity information are integrated into prompt learning as external knowledge to enhance prompt guidance for topic-oriented sarcasm detection.

To assess the EKPL model, we conducted experiments on the *ToSarcasm* dataset. The model proves effective in low-resource small sample environments and also achieves good performance when there are enough training examples. The primary contributions of this paper are as follows:

(1) We propose an entity knowledge-based prompt learning model (EKPL), which effectively utilizes PLM through prompt learning for topic-oriented sarcasm detection.

(2) We integrate entity knowledge and entity context knowledge extracted from the knowledge graph into prompt learning to enhance prompt guidance.

(3) We carry out comprehensive experiments on publicly available datasets, with the experimental results showcasing the effectiveness and superior performance of our proposed model in both low-resource and data-rich scenarios.

The rest of this paper is organized as follows: we introduce related work in section 2. In section 3, we explain our target task. And we introduce the details of EKPL in section 4. Section 5 introduces the experimental content and provides a detailed analysis. In section 6 we conclude and look toward future work.

The preparation of manuscripts which are to be reproduced by photo-offset requires special care. Papers submitted in a technically unsuitable form will be returned for retyping, or canceled if the volume cannot otherwise be finished on time.

2. Related Work

2.1. Sarcasm detection

In recent years, scholars have increasingly focused on sarcasm detection in social media comments. Sarcasm detection for social media comments is a challenging task, which is related to multiple factors such as the authenticity of the comment, the author's intention, and the form of text. Following previous works [10-11], we define sarcastic comments as comments that convey sentiments or thoughts opposite to the user's genuine emotions or intentions. The rule-based and dictionary-based method requires researchers to manually design language rules [12] or extract simple statistical patterns from the corpus [13], and then learn through the statistical machine-learning method [14-15] to distinguish the sarcastic tendency of comments.

Thanks to its powerful data-independent modeling capabilities and excellent performance, deep learning technology has recently garnered significant attention in the realm of sarcasm detection. For example, the study [16] proposed a retrieval detection method for verbal sarcasm by retrieving implicit knowledge in the open domain. In [17], Wang et al. proposed a multi-modal mutual learning network for multi-modal sarcasm detection. In addition, transformer-based models have also been applied to sarcasm detection [18-19]. In fact, the determination of sarcasm depends heavily on the topic information. In [20], Lin et al. used the Latent Diliket Allocation (LDA) topic model to build a key feature vocabulary based on consumer topic reviews and discovered sarcastic feature words based on consumer review topics. In response to the problem that topic information cannot be used in multi-label emotion detection tasks, the study [21] proposed a topic-enhanced capsule network to learn the underlying topic information. In [22], Wu et al. performed topic modeling on input images and text and combined latent multi-modal topic features to enrich context.

Topic-oriented sarcasm detection is a type of sarcasm detection, and they all need to detect the sarcastic tendency of sentences. The difference is that the former requires a specific topic to determine whether the sentence is a satirical expression for that topic. This means that the same sentence may have different sarcastic labels when the topic is different. Therefore, compared with traditional sarcasm detection, topic-oriented sarcasm detection is closer to real scenes and more challenging [5]. Inspired by the topic-oriented sarcasm detection task, this paper proposes an Entity Knowledge-based Prompt Learning (EKPL) model. Different from existing methods for this task, we utilize an external knowledge base to link entities in topics and comments, and we introduce entity context information to provide a better explanation of topics and comments.

2.2. Prompt Learning

Existing PLMS have demonstrated their strong performance in various classification tasks, such as ELMo [23], BERT [24], and Roberta [25]. Many studies are also exploring how to better utilize these PLMs. Previous studies are mainly divided into two ways to leverage PLMs: feature-based and fine-tuning. The feature-based approach regards the PLM as a feature extractor, while the fine-tuning approach treats the PLM as an initialized backbone for continued training on downstream tasks [26]. Despite the great success of the above methods, there is an inevitable huge disparity between the language model pre-training and the target task fine-tuning process [27-28].

Fortunately, many studies have explored another reliable paradigm for leveraging PLMs, called prompt learning. Prompt learning can show the potential to further improve performance by narrowing the disparity between language model pre-training and target task training [29-30]. Later, the study [7] proposed the Pattern Exploration Training (PET) method for few-shot learning. This method replaces labels with [MASK] tokens in a manually constructed template, converts the original classification task into a cloze form, and relies on the language model to predict the words filled in at [MASK]. In response to the problem that previous prompt learning-based methods did not consider external information, the study [31] proposed a knowledge-based prompt learning method. This method improves the coverage of labels by integrating external knowledge bases and demonstrates the effectiveness of knowledge tuning in zero-shot and few-shot text classification tasks. In [32], Xie et al. proposed a method to extract prior knowledge from pre-trained language models using prompts. This method enables the model to effectively extract relevant information from a huge language knowledge base for different relationship types by designing specific query instructions. We leverage prompt learning to use Roberta for topic-oriented sarcasm detection, where the template combines entity knowledge information and learnable tokens.

2.3. Knowledge information utilization

In recent years, external knowledge information has attracted widespread attention in several natural language processing tasks [33-36], including sarcasm detection tasks. Some studies assist in judging the sarcastic tendency of a given text by introducing contextual information [37]. From a multimodal perspective, many studies utilize multimodal data such as images and videos to help with sarcasm detection [38-40]. In addition, some researchers have used knowledge graphs to extract information from text for sarcasm detection [10-11]. The good flexibility of knowledge graphs allows them to be used without external information (such as contextual information and image data). Therefore, we use knowledge graphs for topic-oriented sarcasm detection.

The knowledge graph is composed of multiple entity nodes and edges describing entity relationships, which can provide rich auxiliary information for learning the semantic features of the context. Knowledge graphs are widely used in various fields through entity linking [41-43]. In [44], Hu et al. proposed a method to align knowledge representation with text representation by applying entity-linking technology. In [45], Dun et al. utilize attention mechanisms to adjust word representations and entity knowledge representations. In addition, some studies propose integrating knowledge information into prompt learning. In [16], Wen et al. used prompt learning to identify texts with connotative knowledge related to a given text, thereby improving the model's ability to understand text semantics.

Different from the above methods, our EKPL model integrates knowledge information obtained from the knowledge graph into PLM and further enhances the comprehension of comment content by fusing knowledge information sequences to explore its application in topic-oriented sarcasm detection task.

3. Task Modeling

The topic-oriented sarcasm detection task is a binary classification task. Our research goal is to detect the sarcastic tendency of social media comments on specific topics. We formally define this task as follows: The input is a topic-comment text pair, which contains a topic text and a comment text. The topic text $T=\{t_1,t_2,t_3,\cdots,t_m\}$ consists of m words, and the comment text $C=\{c_1,c_2,c_3,\cdots,c_n\}$ consists of n words. Topic-oriented sarcasm detection aims to predict the label $y\in\{0,1\}$ of a comment C on a specific topic T, where 1 represents sarcasm comments and 0 represents non-sarcasm comments.

4. Methodology

In this section, we first elaborate on the method for acquiring knowledge, then introduce the method of leveraging PLM through prompt learning in our topic-oriented sarcasm detection task. Finally, we detail its knowledge enhancement method.

The overall architecture of the EKPL model is shown in Figure 3. The input of EKPL is the topic and comment text, and its output is the predicted probability of the labels. Specifically, we first leverage prompt learning to build a topic-oriented template and add learnable tokens to this template. Entity knowledge and entity context knowledge are then extracted from the knowledge graph and their representations are integrated into the template. Finally, we map the labels of sarcastic comments to corresponding words and predict them through a module called Veralizer.

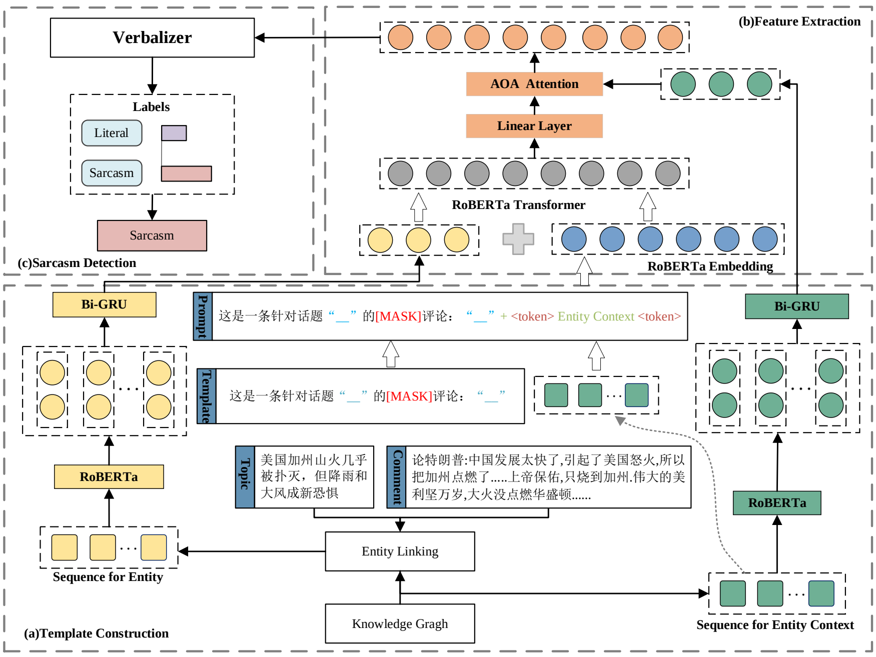

Fig. 3. The overall architecture of the EKPL model

4.1. Knowledge Acquisition

Entity knowledge acquisition. The utilization of external knowledge to enhance the performance of PLM has been extensively researched [46-47]. The purpose of this part is to obtain relevant entity knowledge in topics and comments by knowledge graphs. The model can better comprehend the expression of topic and comment content in this way. Entity linking is the most common method of leveraging knowledge graphs. We use the *TagMe* [48] tool to distinguish entities in topics and comments as shown in Figure 4. Then we align the entities with the corresponding entities in the knowledge graph *Wikidata* [9]. For example, the topic mentions the entity *"特朗普"* (Trump), we link it to and align it with the entity *"唐纳德特朗普"* (Donald Trump) in *Wikidata*. Through this step, we get entity sequence $EN = \{en_1, en_2, en_3, \cdots, en_n\}$.

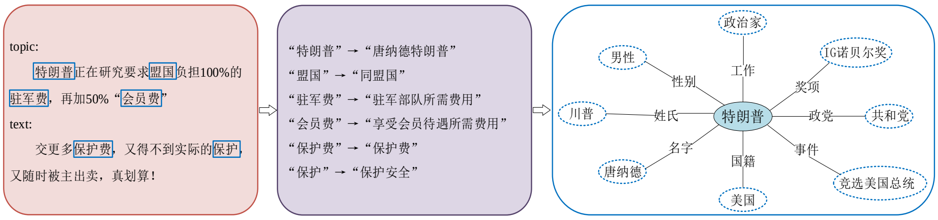

Fig. 4. The process of knowledge extraction

Entity context knowledge acquisition. The entity context is obtained based on the corresponding entities aligned above. For entity en_i, we define the entity context as the entity node adjacent to en_i in the knowledge graph. Therefore, we extract the entity node en_j that is related to the current entity en_i and has a distance of $ec(en_i)$ as the entity context r:

$$ec(en_i) = \left\{ en_i \left| (en_i, r, en_j) \in K \wedge \left(\left| d_{en_j} - d_{en_j} \right| \right) = 1 \right. \right\} \tag{1}$$

where r is the relationship between entity en_i and entity en_j, K is the knowledge graph, and d is the distance between entity en_i and entity en_j. For example, the entity nodes adjacent to the entity *"唐纳德特朗普"* (Donald Trump) are *"美国总统"* (President of the United States), *"IG诺贝尔奖"* (IG Nobel Prize), *"政治家"* (Politician), *"共和党"* (Republican Party) and so on. These adjacent entity nodes form the entity context of entity *"唐纳德特朗普"* (Donald Trump). Then we can get the entity context sequence: $EC = \{ec(en_1), ec(en_2), ec(en_3), \cdots, ec(en_n)\}$.

4.2. Prompt Learning

To address the challenge of data scarcity during fine-tuning, we use a different PLM utilization paradigm, namely prompt learning. Similar to pre-training, prompt learning uses a cloze-style approach to the tuning process. Therefore, it makes more effective use of pre-training information, thereby further improving performance in data-scarce conditions [8]. In this section, we introduce in detail the proposed entity knowledge-based prompt learning model, which uses entity knowledge in prompt learning to complete the topic-oriented sarcasm detection task.

In the topic-oriented sarcasm detection task, the input to our model is a topic-comment text pair that contains a topic text and a comment text. Assume that the topic contains m words and the comment contains n words. We denote the topic text sequence as $T = \{t_1, t_2, t_3, \cdots, t_m\}$ and the comment text sequence as $C = \{c_1, c_2, c_3, \cdots, c_n\}$. To be aligned with the pre-training process, we use a task-related template te to encapsulate the input. This template is carefully designed and summarized based on the characteristics of the topic-oriented sarcasm detection task. We mask a keyword in the template and then add topics and comments to this template. The specific content and representation of the template are as follows:

这是一条针对话题“T”*的*[MASK] *评论*“C”

(This is a [MASK] comment " C " for topic " T ")

$$te = \left[z_1^{te}, \cdots, T, z_i^{te}, \langle mask \rangle, \cdots, z_w^{te}, C \right] \tag{2}$$

Meanwhile, the use of entity information and entity context information containing rich knowledge can effectively promote the topic-oriented sarcasm detection task. Therefore, we propose a method to incorporate entity knowledge representations that are extracted from knowledge graphs into prompt learning. We add the obtained entity context to the template that we constructed as a supplementary explanation of the entities in the template, thereby forming a new interpretable template to advance our prompt learning model.

4.3. Entity knowledge utilization

First, we add the entity context sequence EC obtained from the knowledge graph to the template te to explain the large number of entities contained in the topics and comments in the template, thereby enhancing the guidance performance of the template.

We splice the entity context sequence EC after the template te to form a new prompt p:

这是一条针对话题“T”*的*[MASK] *评论*“C”$| EC$

(This is a [MASK] comment " C " for topic " T "$| EC$)

Inspired by [49], prompt with special learnable tokens can make prompt learning more effective. Therefore, we further add learnable tokens to the prompt p. We insert two special learnable tokens "<ht>" (head token) and "＜tt＞" (tail token) into the head and tail of the entity context sequence EC. The special learnable tokens are randomly initialized and updated during training. The specific content and expression of the prompt are as follows:

这是一条针对话题 “T”*的* [MASK] *评论* “C”|<ht> EC ＜tt＞
(This is a [MASK] comment " C " for topic " T "|<ht> EC ＜tt＞)

$$p=\left[z_1^{te},\cdots,T,z_i^{te},\langle mask\rangle,\cdots,z_w^{te},C\middle|\langle ht\rangle,EC,\langle tt\rangle\right] \tag{3}$$

Then, we divide Roberta's encoder into two parts: embedding and transformer. As shown in formula (4):

$$RoBERTa\_Encoder(p)=RoBERTa\_Transformer\left(RoBERTa\_Embedding(p)\right) \tag{4}$$

We input the prompt p into the embedding layer of Roberta for prompt embedding. As shown in formula (5):

$$e_1^{te},\cdots,e^T,e_i^{te},e_{mask}^{te},\cdots,e_w^{te},e^C\left|e_{ht},e^{EC},e_{tt}\right.=RoBERTa\_Embedding(p) \tag{5}$$

where e_i^{te} is the embedding of the i-th word in template te. e^T, e^C and e^{EC} are the embeddings of topic text sequence, comment text sequence, and entity context sequence respectively. e_{mask}^{te}, e_{ht} and e_{tt} are the embeddings of mask <mask>, learnable tokens <ht> and <tt> respectively.

Next, we encode the entity sequence EN and entity context sequence EC separately. We encode entity sequence EN and entity context sequence EC with Roberta to obtain entity encoding $EN'=\left(r_1^{EN},r_2^{EN},\cdots,r_n^{EN}\right)$ and entity context encoding $EC'=\left(r_1^{EC},r_2^{EC},\cdots,r_n^{EC}\right)$:

$$r_1^{EN},\cdots,r_n^{EN}=RoBERTa\left(en_1,\cdots,en_n\right) \tag{6}$$

$$r_1^{EC},\cdots,r_n^{EC}=RoBERTa\left(ec\left(en_1\right),\cdots,ec\left(en_n\right)\right) \tag{7}$$

We utilize the Bidirectional Gated Recurrent Unit (Bi-GRU) to perform feature extraction on entity encoding EN' and entity context encoding EC'. Bi-GRU utilizes gating mechanisms to capture long-term information of sequences and is less complex than Bidirectional Long Short-Term Memory (Bi-LSTM). The formulas are as follows:

$$h_1^{EN},\cdots,h_n^{EN}=Bi-GRU\left(r_1^{EN},\cdots,r_n^{EN}\right) \tag{8}$$

$$h_1^{EC},\cdots,h_n^{EC}=Bi-GRU\left(r_1^{EC},\cdots,r_n^{EC}\right) \tag{9}$$

where h_i^{EN} and h_i^{EC} are feature vectors containing entity sequence information and entity context sequence information respectively.

Knowledge integration. We choose the head feature h_1^{EN} and tail feature h_n^{EN} in the feature vector $h_1^{EN},\cdots,h_n^{EN}$ that contains long-term and bidirectional entity sequence information output through Bi-GRU as the representation of entity knowledge. These two vectors are then added to the learnable tokens e_{ht} and e_{tt} in the prompt embedding:

$$\hat{e}_{ht}=e_{ht}+h_1^{EN} \tag{10}$$

$$\hat{e}_{tt}=e_{tt}+h_n^{EN} \tag{11}$$

Next, we extract features from the prompt embedding containing entity knowledge to obtain prompt features through Roberta's transformer:

$$l_1^{te},\cdots,l^C\left|l_{ht},\cdots,l_{tt}\right.=RoBERTa\_Transformer\left(e_1^{te},\cdots,e^C\left|\hat{e}_{ht},\cdots,\hat{e}_{tt}\right.\right) \tag{12}$$

After getting the prompt features, we perform a linear transformation on them:

$$L=\left(l_1^{te},\ldots,l^C \mid l_{ht},\ldots,l_{tt}\right)\cdot W^T+b \tag{13}$$

where W is the weight matrix with dimension [256,768], b is the bias vector with dimension [1,256] and L is the output matrix.

To better learn the relevant information between prompt and entity context, we introduce an Attention-over-Attention (AOA) mechanism after the output matrix L. This mechanism can further refine the relationship between hints and entity context by calculating the word pair correlation matrix, thereby improving the performance of the model. Firstly, we calculate an interaction matrix I:

$$I=L\cdot\left(h_i^{EC}\right)^T \tag{14}$$

where $\left(h_i^{EC}\right)^T$ is the transpose of the entity context feature vector.

Each row or column of this interaction matrix I represents the word pair correlation between topic and comment. We perform *softmax* normalization on each row to obtain an attention matrix α of the prompt relative to the entity context. Then we perform *softmax* normalization on each column to obtain an attention matrix β of the entity context relative to the prompt. Then we average β by column to get the EC-level attention representation $\bar{\beta}$, and do the dot multiplication with $\boldsymbol{\alpha}$ to get the prompt-level attention representation γ:

$$\gamma=\alpha\cdot\bar{\beta}^T \tag{15}$$

Finally, the correlation representation s between the prompt and the entity context is the weighted sum of the output matrix L and the attention representation γ:

$$s=L^T\cdot\gamma \tag{16}$$

Veralizer. We use a module called Veralizer to simulate the process of predicting mask words in pre-training. This module maps labels to corresponding words, which are called mapping words. Inspired by [26], we convert each label into words with similar meanings through a similar translation method. For example, the label *“讽刺”* (sarcasm) can be similarly translated as *“反讽”* (irony) and *“讥讽”* (satire). Figure 5 shows the mapping details between the label *“讽刺”* (sarcasm) and its mapping words. As shown in Figure 6, the topic-oriented sarcasm detection is a binary classification task, we set the label opposite to the label *“讽刺”* (sarcasm) as *“字面”* (literal) based on the template content *“这是一条针对话题 “T”的* [MASK] *评论 “C””*.

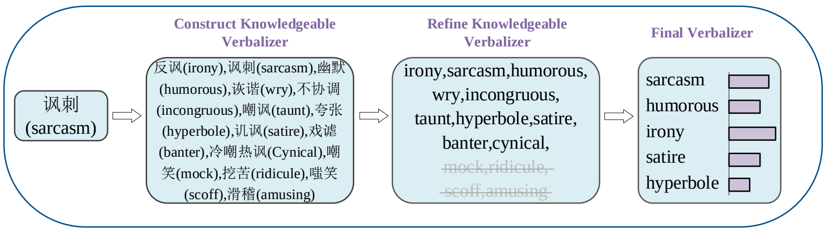

Fig. 5. Mapping words corresponding to the label *“讽刺”* (sarcasm)

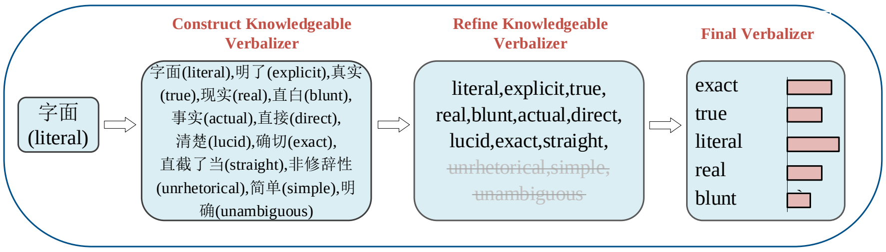

Fig. 6. Mapping words corresponding to the label *“字面”* (literal)

First, we use *Related Words*[‡] and *Google Translation*[§] to obtain words with similar meanings to the labels. *Related Words* is a knowledge graph that integrates various resources, including word embeddings, *ConceptNet*, etc. Then we select 10 mapping words based on the word similarity provided by *Related Words* and *Google Translation*.

Finally, we use MLP to predict the distribution probability of the <mask> position mapping word:

$$P(mw|p) = MLP(s_{mask}^{te}) \tag{17}$$

where mw is the mapping word, $P(mw|p)$ is the probability distribution of the mapping word mw, s_{mask}^{te} is the feature of the mask <mask> in the correlation representation s.

Since each mapping word has a different importance in the corresponding label, we allocate a learnable weight q_{mw} to each mapping word. The label probability $P(y|p)$ can be obtained through the learnable weight q_{mw} and the probability distribution of the mapping word $P(mw|p)$:

$$P(y|p) = sigmoid\left(\sum_{mw \in S_y} q_{mw} P(mw|p)\right) \tag{18}$$

where S_y is the mapping word set corresponding to the label.

We use cross-entropy loss and L2 loss function to train and optimize our EKPL model. In addition, we use dropout to prevent model overfitting, thereby enhancing the model's generalization ability and robustness:

$$Loss = -\sum_{i=1}^{N}\sum_{j=1}^{K} y_i^j \log P(y_i^j|p) + \lambda\|\Theta\|^2 \tag{19}$$

where N is the total number of training samples and K is the total number of categories. y_i^j is the true label of training sample i belonging to category j. $P(y_i^j|p)$ is the probability that training sample i belongs to category j. λ is the L2 regularization coefficient. Θ represents a trainable parameter in the model (e.g. Weight_decay).

[‡] https://relatedwords.org
[§] https://translate.google.com

5. Experiment

In this section, we will verify the validity of the model through experiments. We first present the details of the dataset and related parameter settings. Then, we present and summarize the results of our model on test data. Finally, we analyze the stability and visualization of the model.

Table 1. Dataset information

Label	Train	Validation	Test	Total	Ratio (%)
irony	1464	486	486	2436	50.01
literal	1461	487	487	2435	49.99
total	2925	973	973	4871	100

5.1. Experiment Setup

To verify the effectiveness of our model, we conduct experiments on the publicly available *ToSarcasm* dataset. *ToSarcasm* is a dataset consisting of 4871 topic-comment text pairs, including 707 topics and 4871 comments. These topics and comments come from the *Guanchazhe*[**] website, and are mostly related to political news and humanities and social science. The statistics of the dataset are shown in Table 1.

Due to the scarcity of public Chinese sarcasm detection datasets, *ToSarcasm* is the only existing publicly known topic-oriented sarcasm detection Chinese dataset. Therefore, we set up "Few-shot" and "Full-scale" in the comparison experiment and ablation experiment to simulate low-resource scenario and sufficient-resource scenario.

Few-shot. We randomly select (5, 10, 20, 50, 100) samples from the original training set as the training set to simulate the real low-resource scenario. Then we create the development set of the same size, and the test set still uses the original test set. Considering that the training set and development set with different sample sizes in low-resource scenarios significantly affect model performance, we repeated the above sampling method on 10 random seeds for the experiment. The results of the experiments were averaged after removing the maximum and minimum values.

Full-scale. We directly use the original training set, development set and test set in the *ToSarcasm* dataset. The ratio of the three sets is 6:2:2 and they are all balanced sets.

Model settings. We use Roberta-base [25] as our PLM. The hidden vector dimension of Roberta-base is 768, and the hidden layer dimension of Bi-GRU is 256. We have set the hidden layer size to 200 for MLP. To prevent overfitting, we use dropout to reduce the density of the model, and we set dropout to 0.2. During the training process, if the learning performance of the model does not improve after more than 10 times, the training will end early. We use the *Adam* optimizer [50] to optimize the parameters of the model.

[**] https://www.guancha.cn/

Considering that our goal is to detect sarcastic comments, we set *“讽刺”* (sarcasm) as a positive example and *“字面”* (literal) as a negative example. We use precision (P), recall (R), accuracy (Acc), and F1-score (F1) to evaluate the model's classification performance. The calculation formula for each evaluation indicator is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

$$Precision = \frac{TP}{TP + FP} \tag{21}$$

$$Recall = \frac{TP}{TP + FN} \tag{22}$$

$$F1 - score = \frac{2 \times P \times R}{P + R} \tag{23}$$

where TP represents the number of samples where the predicted label and the actual label both are *“讽刺”* (sarcasm); FP represents the number of samples where the predicted label is *“讽刺”* (sarcasm) but the actual label is *“字面”* (literal); FN represents the number of samples where the predicted label is *“字面”* (literal) but the actual label is *“讽刺”* (sarcasm); TN represents the number of samples where the predicted label and the actual label both are *“字面”* (literal).

5.2. Comparative Methods

In this section, we compare our EKPL model with traditional word vector models (1-3), which do not use PLM. Furthermore, we compare our EKPL model with feature-based models (4-5) and standard fine-tuned models (6-7). We further compare our EKPL model with several other powerful models for prompt learning (8-9). The details are as follows:

(1) **Bi-LSTM** [51]: Use bidirectional LSTM to extract features from sentences and targets respectively, and then concatenate the extracted hidden layer features, which are finally used as classification features for sarcasm detection.

(2) **MIARN** [52]: Effectively captures the contextual semantics of multi-dimensional information through the attention mechanism and Bi-LSTM, and then learns the inconsistency of sarcastic expressions in topics and comments, ultimately achieving accurate sarcasm detection.

(3) **ADGCN** [53]: Utilizes external sentiment knowledge and adaptive dynamic graph convolutional networks to model sentiment inconsistencies in topics and comments to detect sarcastic expressions.

(4) **BERT** [24]: The pre-trained language model BERT-base-Chinese can semantically represent target sentences and capture contextual semantic relationships through pre-training and fine-tuning.

(5) **ADGCN-BERT** [53]: Combines the Adaptive Dynamic Graph Convolution Network (ADGCN) and the pre-trained language model BERT-base-Chinese.

(6) **KL-BERT** [54]: A sarcasm detection model based on BERT and incorporating common sense knowledge.

(7) **KC-ISA-BERT** [55]: An implicit sentiment analysis model that uses a joint attention mechanism to integrate external common sense and contextual features.

(8) **PET** [7]: Use the "[CLS]s[SEP]t[SEP] this is [MASK]" template as the input of the pre-trained language model BERT-base-Chinese, and predict whether the [MASK] position corresponds to "sarcasm" or "non-sarcasm"

(9) **TOSPrompt** [5]: Prompt learning is used to construct a topic-oriented sarcastic expression prompt template to determine the sarcastic tendency of comments on specific topics.

5.3. Experimental Results and Analysis

Our experiments on the *ToSarcasm* dataset are shown in Table 2 and Table 3. From Table 2 and Table 3 we can derive the following results:

Table 2. Comparison between our model and other models in **Few-shot**. The best results among the comparison models are marked with "*"

Model	Few-shot									
	5		10		20		50		100	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Bi-LSTM	23.17	26.06	28.27	30.76	35.56	37.58	31.60	33.36	39.67	41.43
MIARN	24.89	26.54	26.23	27.76	19.52	20.45	18.45	20.84	40.67	42.25
ADGCN	27.53	29.30	30.09	31.98	31.91	33.56	34.03	35.94	41.90	43.57
BERT	35.24	36.75	33.71	36.32	35.90	37.70	38.57	40.38	47.23	50.65
KL-BERT	35.89	36.27	34.66	36.71	35.41	37.83	38.72	40.59	48.17	50.90
ADGCN-BERT	30.94	32.75	36.04	38.61	42.38	44.97	50.57	52.30	58.16	60.98
KC-ISA-BERT	29.06	30.97	36.75	38.32	42.88	44.80	50.79	51.80	58.23	61.23
PET	36.67	37.80	37.07	38.78	43.44*	45.20	50.08	52.63	59.87	61.57
TOSPrompt	37.52*	39.61*	41.67*	43.22*	43.02	45.68*	51.16*	53.39*	60.22*	62.88*
EKPL	40.83	41.25	42.77	43.92	46.40	48.34	52.52	54.60	61.29	63.51

Comparison with Word Vector Models. First, we analyze the results between traditional word vector models and our entity knowledge-based prompt learning model. We used Chinese Word Vectors [56] to initialize word vectors for the three comparison models of Bi-LSTM, MIARN, and ADGCN. According to the experimental results, our entity knowledge-based prompt learning model has significantly improved the four evaluation indicators of precision, recall, accuracy, and F1-score compared to the traditional word vector models. Because the traditional word vector models have shortcomings in text semantic representation. In addition, the PLM-based models have also improved their indicators compared with the traditional word vector models. It can be shown that in the topic-oriented sarcasm detection task, the PLM-based models have better results.

Table 3. Comparison between our model and other models in **Full-scale**. The best results among the comparison models are marked with "*"

Model	Full-scale	
	Acc	F1
Bi-LSTM	63.72	66.65
MIARN	65.32	68.25
ADGCN	65.90	69.19
BERT	69.17	69.09
KL-BERT	69.66	70.14
ADGCN-BERT	70.40	70.83
KC-ISA-BERT	70.57	71.88
PET	70.70	71.44
TOSPrompt	71.76*	73.20*
EKPL	73.77	74.04

Comparison with other PLM utilization models. Second, we compare our prompt learning model with feature and fine-tuning based PLM utilization models (KL-BERT, BERT, KC-ISA-BERT, ADGCN-BERT). It can be seen that our prompt learning model achieves better performance. It proves that our prompt learning model can better utilize PLM information. In addition, through the comparison of the above four PLM utilization models and prompt learning models (PET, TOSPrompt), we found that prompt learning has better results in topic-oriented sarcasm detection tasks. A simple sentence-level sarcasm detection model cannot complete the topic-oriented sarcasm detection task well, but this problem can be better solved with the help of the topic information prompt.

Comparison with conventional prompt learning models. Finally, we validate the contribution of external knowledge to prompt learning. We compared our entity knowledge-based prompt learning model with conventional prompt learning models (PET, TOSPrompt). Compared with PET, our EKPL model improves accuracy and F1-score by 2.01% and 0.84% respectively. Compared with TOSPrompt, our EKPL model improves accuracy and F1-score by 3.07% and 2.60% respectively. This shows that external knowledge can effectively guide prompt learning, thereby improving model performance.

The above results show that our EKPL model achieves the best classification performance when the training data is sufficient. Our EKPL model still has good performance when training data is insufficient. Therefore, this model is suitable for topic-oriented sarcasm detection tasks in real scenarios. We summarize the reasons for the good performance of the EKPL model into the following two points: 1) EKPL uses prompt learning to model topic-oriented sarcasm detection tasks, and by designing a topic-oriented sarcastic expression prompt learning template, it can better learn the sarcastic expression information in comments. 2) EKPL uses entity knowledge and entity context knowledge to eliminate ambiguity caused by entities in topics and comments. It uses the attention mechanism to effectively integrate entity information into prompt learning.

Comparison with EKPL variants. We conducted ablation experiments to verify the effectiveness of each part of the EKPL model. Variants of EKPL are as follows:

EKPL-EC: EKPL-EC is a variant of EKPL without entity context knowledge. The specific method is to remove the entity context sequence in the prompt and remove the AOA mechanism and the entity context features input to the AOA mechanism.

EKPL-EN: EKPL-EN is a variant of EKPL without entity knowledge. We remove learnable tokens so that entity sequence features cannot be added to the corresponding positions of the learnable tokens in the prompt embedding.

EKPL-EC-EN: EKPL-EC-EN is a variant of EKPL without entity knowledge and entity context knowledge. This variant only uses a manually designed template for prompt learning and is equivalent to the regular prompt learning model.

EKPL-LW: EKPL-LW is a variant of EKPL without learnable weight for mapping words.

EKPL-PT: EKPL-PT is a variant of EKPL that does not use prompt learning. We feed topics and comments directly into the model and keep only two learnable tokens.

EKPL-TM: EKPL-TM is a variant of EKPL without template. We do not use templates for prompt learning, only keeping <mask> and entity knowledge injection.

EKPL-AOA: EKPL-AOA is a variant of EKPL without the AOA mechanism.

Table 4. Ablation study in **Few-shot**

variants	Few-shot									
	5		10		20		50		100	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
-PT	32.14	33.49	33.98	34.21	33.90	34.00	35.61	36.37	37.78	38.82
-TM	32.85	33.76	32.47	33.95	34.35	35.49	35.69	36.21	39.21	40.21
-EC-EN	35.25	36.30	35.92	36.53	37.34	38.57	38.60	39.08	43.51	44.73
-EN	37.80	38.54	38.78	39.59	44.20	45.03	50.63	51.41	59.73	60.51
-EC	37.48	38.74	39.57	40.14	45.37	46.07	51.64	52.16	60.43	61.17
-LW	38.12	39.68	40.50	41.16	45.69	46.90	51.41	52.20	60.74	61.90
-AOA	39.51	40.17	41.48	42.43	46.39	47.53	52.12.	53.37	61.29	62.10
EKPL	40.83	41.25	42.77	43.92	46.40	48.34	52.52	54.60	61.96	63.51

As shown in Table 4 and Table 5, the variants of EKPL show varying degrees of performance degradation, which indicates that each part of EKPL is effective. Variant EKPL-PT and variant EKPL-TM have the most significant performance degradation. One possible reason is that after losing the guidance of the prompt, it is difficult for the model to understand the comment content based on the topic, resulting in more errors when predicting the meaning of the <mask> position. The result proves the importance of prompt learning and the construction of a template for prompt learning. When we remove entity context knowledge, the Acc and F1 of EKPL-EC decrease by 2.40% and 1.28% respectively. This shows that comprehensive entity context knowledge is helpful for understanding entities in topics and comments. When we ignore entity knowledge, the Acc and F1 of EKPL-EN decrease by 2.54% and 1.39% compared with EKPL. The result shows that entity knowledge plays an important role in entity disambiguation. It also provides a basis for effectively combining entity context knowledge. When we remove external knowledge from EKPL, the indicators of EKPL-EC-EN decrease. For example, there is a 2.29% drop compared to EKPL in terms of F1-score. The research result shows that entity knowledge is instructive for EKPL to detect sarcasm. Furthermore, we observe that the variant EKPL-LW performs comparably to fine-tuning without learnable weight for mapped words. Probably because PLM has a fixed prior preference for [MASK] prediction, which is determined by the pre-training process. The learnable weights of mapping words map the mapping words to the label through weighted summation to obtain more accurate results.

Table 5. Ablation study in Full-scale

variants	Full-scale	
	Acc	F1
-PT	68.48	69.06
-TM	69.56	70.49
-EC-EN	70.44	71.75
-EN	71.23	72.65
-EC	71.37	72.76
-LW	72.18	73.30
-AOA	72.32	73.93
EKPL	73.77	74.04

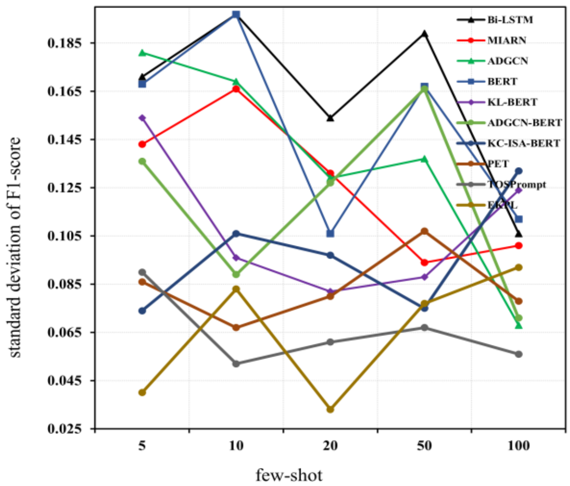

Fig. 7. The standard deviation of F1-score in few-shot settings

5.4. Stability Analysis

The stability of a model is also an important indicator for evaluating model performance, so we measured and analyzed the stability of our EKPL model. Model stability is measured as the standard deviation of F1-score. As shown in Figure 7, we compared the standard deviations of the 10 experimental results of each model under few-shot in the comparative experiment. It can be found that our EKPL model is more stable than feature-based, fine-tuning and other prompt learning models in most cases through comparison. This proves that our model has great stability.

In addition, we found that the standard deviation of the model will show a downward trend as the few-shot increases. We noticed that the word vector-based models tend to be more stable than various PLM utilization models in the case of smaller few-shots. It may be that PLM-based models are prone to overfitting when there is insufficient data, resulting in poor generalization ability and high randomness of the model.

5.5. Visualization of mapping word weights

We analyze the weight of each answer word after training. We average the answer weights under few-shot settings.

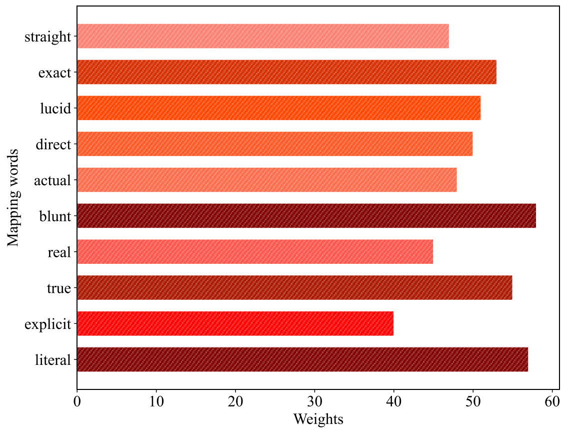

Fig. 8. The mapping word weights of literal label

Among the mapping words corresponding to the label *“字面”* (literal), *“直白”* (blunt) has the highest weight, while *“明了”* (explicit) has the smallest weight. As shown in Figure 8, it may be that PLM has a strong preference for the word *“直白”* (blunt). This result suggests that PLM's perception and understanding of words may be different from humans. As shown in Figure 9, we performed a visual analysis of the weight of each mapped word after training. Among the mapping words corresponding to the label *“讽刺”* (sarcasm), *“讽刺”* (sarcasm) and *“反讽”* (irony) have higher weights, which is in line with our predictions.

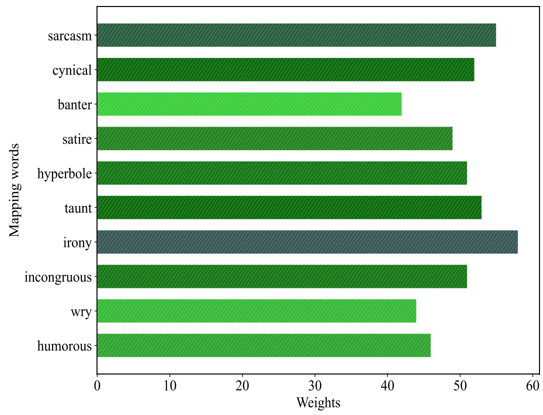

Fig. 9. The mapping word weights of sarcasm label

6. Conclusion

In this paper, we propose an entity knowledge-based prompt learning model (EKPL) to solve the topic-oriented sarcasm detection task. The core idea of the model is to use prompt learning to guide PLM. It adopts the same goal as the language model pre-training, significantly narrowing the gap between pre-training and target task training. In addition, we inject entity knowledge into our well-designed template to refine the prompt representation. This can further enhance our sarcasm detection model. To the best of our knowledge, this is the first work to combine prompt learning and knowledge graphs for topic-oriented sarcasm detection.

To evaluate our model, we conduct experiments on the publicly available *ToSarcasm* dataset. Comparison with different PLM utilization methods and other prompt learning models fully illustrates the superiority of our prompt learning method and the effectiveness of external knowledge on our prompt learning model. Furthermore, we conduct a detailed discussion and analysis of the results for a comprehensive understanding of our model.

In future work, we will look for better knowledge representations and incorporate them as explicit features into prompt learning and deep neural networks to further improve the performance of sarcasm detection models.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant NO.62076006), the Opening Foundation of the State Key Laboratory of Cognitive Intelligence, iFLYTEK (Grant NO.COGOS-2023HE02), and the University Synergy Innovation Program of Anhui Province (Grant NO.GXXT-2021-008).

References

1. Kanakam, R., & Nayak, R. K. (2021, June). Sarcasm Detection on Social Networks using Machine Learning Algorithms: A Systematic Review. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1130-1137). IEEE.
2. Zhang, S., Zhao, T., Wu, H., Zhu, G., & Li, K. (2023). TS-GCN: Aspect-level sentiment classification model for consumer reviews. Computer Science and Information Systems, 20(1), 117-136.
3. Khan, M., & Malviya, A. (2020, February). Big data approach for sentiment analysis of twitter data using Hadoop framework and deep learning. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (pp. 1-5). IEEE.
4. Wang Z, Hu S, Liu W. Product feature sentiment analysis based on GRU-CAP considering Chinese sarcasm recognition[J]. Expert Systems with Applications, 2024, 241: 122512.
5. Liang, B., Lin, Z., Qin, B., & Xu, R. (2022, October). Topic-Oriented Sarcasm Detection: New Task, New Dataset and New Method. In Proceedings of the 21st Chinese National Conference on Computational Linguistics (pp. 557-568).
6. Feng K, Huang L, Wang K, et al. Prompt-based learning framework for zero-shot cross-lingual text classification[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108481.
7. Schick, T., & Schütze, H. (2021, April). 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).
8. Schick, T., & Schütze, H. (2021, June). It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 2339-2352).
9. Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10), 78-85.
10. Wen, Z., Gui, L., Wang, Q., Guo, M., Yu, X., Du, J., & Xu, R. (2022). Sememe knowledge and auxiliary information enhanced approach for sarcasm detection. Information Processing & Management, 59(3), 102883.
11. Ren, Y., Wang, Z., Peng, Q., & Ji, D. (2023). A knowledge-augmented neural network model for sarcasm detection. Information Processing & Management, 60(6), 103521.
12. Pavan Kumar, M. R., & Jayagopal, P. (2023). Context-sensitive lexicon for imbalanced text sentiment classification using bidirectional LSTM. Journal of Intelligent Manufacturing, 34(5), 2123-2132.
13. Ortega-Bueno, R., Rosso, P., & Pagola, J. E. M. (2022). Multi-view informed attention-based model for Irony and Satire detection in Spanish variants. Knowledge-Based Systems, 235, 107597.
14. Vinoth, D., & Prabhavathy, P. (2022). An intelligent machine learning-based sarcasm detection and classification model on social networks. The Journal of Supercomputing, 78(8), 10575-10594.
15. Govindan, V., & Balakrishnan, V. (2022). A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection. Journal of King Saud University-Computer and Information Sciences, 34(8), 5110-5120.
16. Wen, Z., Wang, R., Chen, S., Wang, Q., Ding, K., Liang, B., & Xu, R. (2023). RDVI: A Retrieval–Detection Framework for Verbal Irony Detection. Electronics, 12(12), 2673.
17. Wang J, Yang Y, Jiang Y, et al. Cross-modal incongruity aligning and collaborating for multi-modal sarcasm detection[J]. Information Fusion, 2024, 103: 102132.

18. Mohan, A., Nair, A. M., Jayakumar, B., & Muraleedharan, S. (2023). Sarcasm Detection Using Bidirectional Encoder Representations from Transformers and Graph Convolutional Networks. Procedia Computer Science, 218, 93-102.
19. Lu Q, Long Y, Sun X, et al. Fact-sentiment incongruity combination network for multimodal sarcasm detection[J]. Information Fusion, 2024, 104: 102203.
20. Lin, H. C. K., Wang, T. H., Lin, G. C., Cheng, S. C., Chen, H. R., & Huang, Y. M. (2020). Applying sentiment analysis to automatically classify consumer comments concerning marketing 4Cs aspects. Applied Soft Computing, 97, 106755.
21. Fei, H., Ji, D., Zhang, Y., & Ren, Y. (2020). Topic-enhanced capsule network for multi-label emotion classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 1839-1848.
22. Wu, S., Fei, H., Cao, Y., Bing, L., & Chua, T. S. (2023). Information Screening whilst Exploiting! Multimodal Relation Extraction with Feature Denoising and Multimodal Topic Modeling. arXiv preprint arXiv:2305.11719.
23. Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: human language technologies, Volume 1 (Long Papers) (pp. 2227–2237).
24. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171–4186).
25. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
26. Jiang, G., Liu, S., Zhao, Y., Sun, Y., & Zhang, M. (2022). Fake news detection via knowledgeable prompt learning. Information Processing & Management, 59(5), 103029.
27. Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., & Liu, Q. (2019). ERNIE: Enhanced language representation with informative entities. arXiv preprint arXiv:1905.07129.
28. Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., & Wang, P. (2020, April). K-bert: Enabling language representation with knowledge graph. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 03, pp. 2901-2908).
29. Liu, X., & Zheng, Y. (2021). Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2021. Gpt understands, too. arXiv preprint arXiv:2103.10385, 6(2).
30. Han, X., Zhao, W., Ding, N., Liu, Z., & Sun, M. (2022). Ptr: Prompt tuning with rules for text classification. AI Open, 3, 182-192.
31. Hu, S., Ding, N., Wang, H., Liu, Z., Wang, J., Li, J., ... & Sun, M. (2021). Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification. arXiv preprint arXiv:2108.02035.
32. Xie, S., Pan, Q., Wang, X., Luo, X., & Sugumaran, V. (2024). Combining prompt learning with contextual semantics for inductive relation prediction. Expert Systems with Applications, 238, 121669.
33. Zhu, C., Chen, M., Zhang, S., Sun, C., Liang, H., Liu, Y., & Chen, J. (2023). SKEAFN: Sentiment knowledge enhanced attention fusion network for multimodal sentiment analysis. Information Fusion, 100, 101958.
34. Jain, P. K., Quamer, W., Saravanan, V., & Pamula, R. (2023). Employing BERT-DCNN with sentic knowledge base for social media sentiment analysis. Journal of Ambient Intelligence and Humanized Computing, 14(8), 10417-10429.
35. Zhang, B., Hu, Y., Xu, D., Li, M., & Li, M. (2022). SKG-Learning: a deep learning model for sentiment knowledge graph construction in social networks. Neural Computing and Applications, 34(13), 11015-11034.
36. He, Z., Wang, H., & Zhang, X. (2023). Multi-Task Learning Model Based on BERT and Knowledge Graph for Aspect-Based Sentiment Analysis. Electronics, 12(3), 737.

37. Yu, Z., Jin, D., Wang, X., Li, Y., Wang, L., & Dang, J. (2023, August). Commonsense knowledge enhanced sentiment dependency graph for sarcasm detection. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence pp. 2423-2431.
38. Chauhan, D. S., Singh, G. V., Arora, A., Ekbal, A., & Bhattacharyya, P. (2022). An emoji-aware multitask framework for multimodal sarcasm detection. Knowledge-Based Systems, 257, 109924.
39. Bedi, M., Kumar, S., Akhtar, M. S., & Chakraborty, T. (2021). Multi-modal sarcasm detection and humor classification in code-mixed conversations. IEEE Transactions on Affective Computing.
40. Yue, T., Mao, R., Wang, H., Hu, Z., & Cambria, E. (2023). KnowleNet: Knowledge fusion network for multimodal sarcasm detection. Information Fusion, 100, 101921.
41. Cheng, L., Wu, D., Bing, L., Zhang, Y., Jie, Z., Lu, W., & Si, L. (2020, November). ENT-DESC: Entity Description Generation by Exploring Knowledge Graph. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) pp. 1187-1197.
42. Li, Y., Zou, B., Li, Z., Aw, A., Hong, Y., & Zhu, Q. (2021, November). Winnowing knowledge for multi-choice question answering. In Findings of the association for computational linguistics: EMNLP 2021 pp. 1157-1165.
43. Wei, X., Huang, H., Nie, L., Zhang, H., Mao, X. L., & Chua, T. S. (2016). I know what you want to express: sentence element inference by incorporating external knowledge base. IEEE Transactions on Knowledge and Data Engineering, 29(2), 344-358.
44. Hu, L., Yang, T., Zhang, L., Zhong, W., Tang, D., Shi, C., ... & Zhou, M. (2021, August). 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) pp. 754-763.
45. Dun, Y., Tu, K., Chen, C., Hou, C., & Yuan, X. (2021, May). Kan: Knowledge-aware attention network for fake news detection. In Proceedings of the AAAI conference on artificial intelligence Vol. 35, No. 1, pp. 81-89.
46. Guan, J., Huang, F., Zhao, Z., Zhu, X., & Huang, M. (2020). A knowledge-enhanced pretraining model for commonsense story generation. Transactions of the Association for Computational Linguistics, 8, 93-108.
47. Yang, A., Wang, Q., Liu, J., Liu, K., Lyu, Y., Wu, H., ... & Li, S. (2019, July). Enhancing pre-trained language representations with rich knowledge for machine reading comprehension. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics pp. 2346-2357.
48. Ferragina, P., & Scaiella, U. (2010, October). 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.
49. Hambardzumyan, K., Khachatrian, H., & May, J. (2021, August). WARP: Word-level Adversarial ReProgramming. 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. 4921-4933.
50. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
51. Graves, A., & Graves, A. (2012). Long short-term memory. Supervised sequence labelling with recurrent neural networks, 37-45.
52. Tay, Y., Luu, A. T., Hui, S. C., & Su, J. (2018, July). Reasoning with Sarcasm by Reading In-Between. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) pp. 1010-1020.

53. Lou, C., Liang, B., Gui, L., He, Y., Dang, Y., & Xu, R. (2021, July). Affective dependency graph for sarcasm detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval pp. 1844-1849.
54. Li, J., Pan, H., Lin, Z., Fu, P., & Wang, W. (2021). Sarcasm detection with commonsense knowledge. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 3192-3201.
55. Xu, M., Wang, D., Feng, S., Yang, Z., & Zhang, Y. (2022, October). Kc-isa: An implicit sentiment analysis model combining knowledge enhancement and context features. In Proceedings of the 29th International Conference on Computational Linguistics pp. 6906-6915.
56. Li, S., Zhao, Z., Hu, R., Li, W., Liu, T., & Du, X. (2018, July). Analogical Reasoning on Chinese Morphological and Semantic Relations. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) pp. 138-143.

Yuhao Zhou received his Bachelor’s degree from the School of Computing Science and Engineering, Anhui University of Science and Technology in 2022, China. At present, he is a MS degree majoring in Computer Science and Technology from the School of Computing Science and Engineering, Anhui University of Science and Technology. His current research interests include natural language processing, sentiment analysis, and complex network.

Shunxiang Zhang received the Ph.D. degree from the School of Computing Engineering and Science, Shanghai University, Shanghai, China, in 2012. He is a Professor at the Anhui University of Science and Technology, Huainan, China. His current research interests include web mining, semantic search, and complex network.

Caiqing Wang received his Bachelor’s degree from the School of Computing Science and Engineering, Anhui University of Science and Technology in 2022, China. At present, he is a MS degree majoring in Computer Science and Technology from the School of Computing Science and Engineering, Anhui University of Science and Technology. His current research interests include natural language processing, web intelligent information processing, and calculation theory.

Yanhui Wang received her Bachelor’s degree from the School of Computing Science and Engineering, Anhui University of Science and Technology in 2022, China. At present, she is a MS degree majoring in Computer Science and Technology from the School of Computing Science and Engineering, Anhui University of Science and Technology. Her current research interests include natural language processing, web intelligent information processing, and sentiment analysis.

Xiaolong Wang received his Bachelor’s degree from the School of Computing Science and Engineering, Anhui University of Science and Technology in 2022, China. At present, he is a MS degree majoring in Computer Technology from the School of Computing Science and Engineering, Anhui University of Science and Technology. His current research interests include natural language processing, web intelligent information processing, and semantic search.

Kuan-Ching Li is currently a Distinguished Professor at Providence University, where he also serves as the Director of the High-Performance Computing and Networking Center. He published more than 450 scientific papers and articles and is the coauthor or coeditor of more than 50 books published by well-known publishers. He is the Editor-in-Chief of IJCSE and IJES and serves as an associate editor for several leading journals. His research interests include parallel and distributed computing, big data, and emerging technologies. He is a fellow of the IET and a Senior Member of the IEEE.

Received: March 20, 2024; Accepted: September 22, 2024.