





Sentence-Level Event Detection Without Triggers via Prompt Learning and Machine Reading Comprehension

Tongtao Ling, Lei Chen^(✉), Huangxu Sheng, Zicheng Cai, and Hai-Lin Liu^{}

Guangdong University of Technology, Guangzhou, China
chenlei3@gdut.edu.cn

Abstract. Sentence-level event detection has traditionally been carried out in two key steps: trigger identification and trigger classification. The trigger words first are identified from sentences and then utilized to categorize event types. However, this classification hugely relies on a substantial amount of annotated trigger words along with the accuracy of the trigger identification process. This annotation of trigger words is labor-intensive and time-consuming in real-world environments. As a solution to this, we propose a model that does not require any triggers for event detection. This model reformulates event detection into a two-tower model that uses machine learning comprehension and prompt learning. Compared to the existing methods, which are either trigger-based or trigger-free, experimental studies on two benchmark event detection datasets (ACE2005 and MAVEN) reveal that our proposed method can achieve competitive performance.

Keywords: Event detection · Prompt learning · Machine reading comprehension

1 Introduction

Information extraction (IE) is an important application of Natural Language Processing (NLP). Event detection (ED) is a fundamental part of IE, aiming at identifying trigger words and classifying event types, which could be divided into two sub-tasks: trigger identification and trigger classification [1]. For example, consider the following sentence “*To **assist** in managing the vessel traffic, Chodkiewicz **hired** a few sailors, mainly Livonian*”. The trigger words are “assist” and “hired”, the trigger-based event detection model is used to locate the position of the trigger words and classify them into the corresponding event types, *Assistance* and *Employment* respectively.

Contemporary mainstream studies on ED concentrate on trigger-based methods. These methods involve initially identifying the triggers and then categorizing the types of events [2–4]. This approach changes the ED task into a multi-stage classification issue, with the outcome of trigger identification also impacting

the categorization of triggers. Therefore, it is crucial to identify trigger words correctly, which requires datasets containing multiple annotated trigger words and event types [5]. However, it is time-consuming to annotate trigger words in a real scenario, especially in a long sentence. Due to the expensive annotation of the corpus, the application of existing ED approaches is greatly limited. It should be noted that trigger words are considered an extra supplement for trigger classification, but event triggers may not be essential for ED [6].

From a problem-solving perspective, ED aims to categorize the type of events and therefore triggers can be seen as an intermediate result of this task [6]. To alleviate manual effort, we aim to explore how to detect events without triggers. Event detection can be considered a text classification problem if the event triggers are missing. But three challenges should be solved: (1) **Multi-label problem**: since a sentence can contain multiple events or no events at all, which is called a multi-label text classification problem in NLP. (2) **Insufficient event information**: triggers are important and helpful for ED [2, 7]. Without trigger words, the ED model may lack sufficient information to detect the event type, and we need to find other ways to enrich the sentence semantic information and learn the correlation between the input sentence and the corresponding event type. (3) **Imbalance Data Distribution**: the data distribution in the real world is long-tail, which means that most event types have only a small number of instances and many sentences may not have events occurring. The goal of ED is also to evaluate its ability in the long-tail scenario.

To detect events without triggers and solve these problems, we propose a two-tower model via machine reading comprehension (MRC) [8] and prompt learning [9]. Figure 1 illustrates the structure of our proposed model with two parts: reading comprehension encoder (RCE) and event type classifier (ETC). In the first-tower, we employ BERT [10] as backbone, and the input sentence concatenates with all event tokens are fed into BERT simultaneously¹. Such a way is inspired by the MRC task, extracting event types is formalized as extracting answer position for the given sequence of event type tokens. In other words, the input sentences are deemed as “Question” and the sequence of event type tokens deemed as “Answer”. This way allows BERT to automatically learn semantic relations between the input sentences and event tokens through self-attention mechanism [11]. In the second-tower, we use the same backbone of RCE and utilize prompt learning methods to predict event types. Specifically, when adding the prompt “This sentence describes a [MASK] event” after the original sentence, this prompt can be viewed as a cloze-style question and the answer is related to the target event type. Therefore, ETC aims to fill the [MASK] token and can output the scores for each vocabulary token. We only use event type tokens in vocabulary and predict event types that score higher than the $\langle none \rangle$ event type. In the inference time, only when these two-tower models predict results are correct can they be used as the final correct answer. In our

¹ For example, we convert event token *employment* to “ $\langle employment \rangle$ ” and add it to vocabulary. All events operate like this. In addition, we add a special token “ $\langle none \rangle$ ” that no events have occurred.

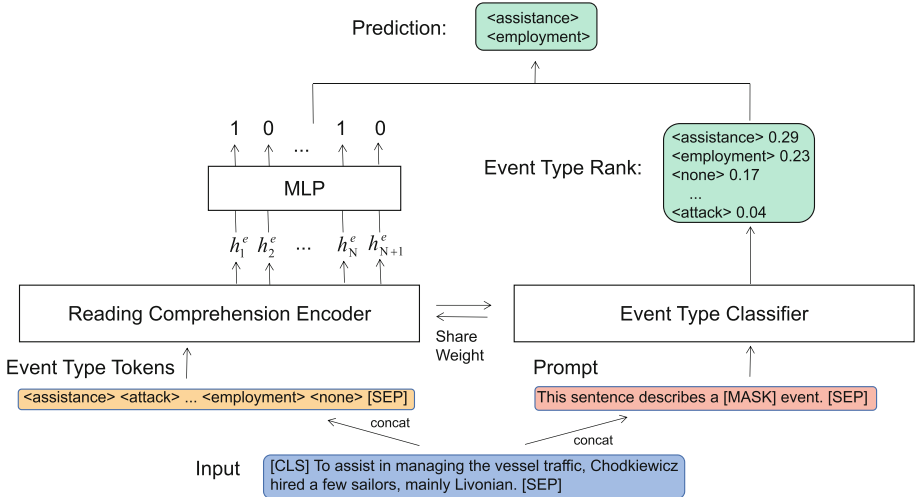


Fig. 1. Overview of our proposed EDPRC. It consists of two modules: reading comprehension encoder (RCE) and event type classifier (ETC).

example from Fig. 1, RCE can predict the answer tokens are $\langle assistance \rangle$ and $\langle employment \rangle$ respectively. In addition, since $\langle assistance \rangle$ and $\langle employment \rangle$ both have higher values than $\langle none \rangle$, we predict *Assistance* and *Employment* as the event type in this sentence.

In summary, we propose a two-tower model to solve the ED task without triggers and call our model **EDPRC: Event Detection via Prompt learning and machine Reading Comprehension**. The main contributions of our work are: (1) We propose a trigger-free event detection method based on prompt learning and machine reading comprehension that does not require triggers. The machine reading comprehension method can capture the semantic relations between sentence and event tokens. The prompt learning method can evaluate the scores of all event tokens in vocabulary; (2) Our experiments can achieve competitive results compared with other trigger-based methods and outperform other trigger-free baselines on ACE2005 and MAVEN; (3) Further analysis of attention weight also indicates that our trigger-free model can identify the relation between input sentences and events, and appropriate prompts in a specific topic can guide pre-trained language models to predict correct events.

2 Related Work

2.1 Sentence-Level Event Detection

Conventional sentence-level event detection models based on pattern matching methods mainly utilize syntax trees or regular expressions [12]. These pattern-matching methods largely rely on the expression form of text to recognize triggers

and classify them into event types in sentences, which fails to learn in-depth features from plain text that contains complex semantic relations. With the rapid development of deep learning, most ED models are based on artificial neural networks such as convolutional neural networks (CNN) [2], recurrent neural network (RNN) [3], graph neural network (GNN) [13] and transformer network [14], and other pre-trained language models [10, 15].

2.2 Machine Reading Comprehension

Machine reading comprehension (MRC) is a difficult task in natural language processing (NLP) that involves extracting relevant information from a passage to answer a question. The process can be broken down into two parts: identifying the start and end points of the answer within the passage [16, 17]. Recently, researchers have been exploring ways to adapt event extraction techniques for use in MRC question answering. One approach is to convert event extraction into a MRC task, where questions are generated based on event schemas and answers are retrieved accordingly [18]. Another approach is to utilize a mechanism like DRC, which employs self-attention to understand the relationships between context and events, allowing for more accurate answer retrieval [19].

2.3 Prompt Learning

In recent years, there has been significant progress in natural language processing (NLP) tasks using prompt-based methods [9]. Unlike traditional model fine-tuning, prompt-tuning involves adding prompts to the raw input to extract knowledge from pre-trained language models like BERT [10] and GPT3 [20]. This new approach allows for the creation of tailored prompts for specific downstream tasks such as text classification, relation extraction, and text generation. By doing so, it bridges the gap between pre-trained tasks and downstream tasks, reducing training time significantly [21]. Additionally, prompt-based learning enables pre-trained language models to gain prior knowledge of a particular downstream task, ultimately improving performance [22].

3 Methodology

In this section, we present the proposed EDPRC in detail for sentence-level event detection without triggers.

3.1 Problem Description

Formally, denote \mathcal{X} , \mathcal{Y} as the sentence set and the event type set, respectively. $\mathcal{X} = \{x_i | i \in [1, M]\}$ contains M sentences, and each sentence x_i in \mathcal{S} is a token sequence $x_i = (w_1, w_2, \dots, w_L)$ with maximum length L . In sentence-level event detection, given a sentence x_i and its ground-truth $y_i \in \mathcal{Y}$, $\mathcal{Y} = \{e_1, e_2, \dots, e_N\}$, we need to detect the corresponding event types for each instance. For sentences

where no event occurred, we add a special token “ $\langle None \rangle$ ” as their event type. This problem can be reformulated as a multi-label classification task with $N + 1$ event types.

3.2 Reading Comprehension Encoder

Inspired by the MRC task, we employ BERT as backbone to design a reading comprehension encoder due to its capability in learning contextual representations of the input sequence. We describe it as follows:

$$Input = [\text{CLS}] \text{ Sentence } [\text{SEP}] \text{ Events} \quad (1)$$

where **Sentence** is the input sentence and **Events** is the event type set (also including “ $\langle None \rangle$ ”). [CLS] and [SEP] stand for the start token and separator token in BERT, respectively. For some event types such as “Business:Lay off” fails to map to a single token according to the vocabulary. In this case, we employ an angle bracket around each event type and remove the prefix, e.g., the event type of “Business:Lay off” is converted to a lower-case “ $\langle lay_off \rangle$ ”. Then, we add $N + 1$ event tokens to the vocabulary and randomly initialize its embeddings. Our objective is to utilize BERT for understanding the correlation between the event types and input sentence, producing accurate representations of event tokens.

After that, we get the token representations by using BERT:

$$h_{[\text{CLS}]}, h_1^w, \dots, h_L^w, h_{[\text{SEP}]}, h_1^e, \dots, h_N^e, h_{N+1}^e = BERT(Input) \quad (2)$$

where h_i^w is the hidden state of the i -th input token. This setup is close to MRC that chooses the correct option to answer question “What happened in the sentence?”. Unlike traditional fine-tuning methods that utilize the [CLS] token to complete classification, we use the hidden states of event tokens to predict the probability of each token being the correct answer. The representation of event tokens:

$$E = h_1^e, \dots, h_N^e, h_{N+1}^e \quad (3)$$

where $E \in \mathbb{R}^{N \times D}$, D is the dimension of token representation. The probability of each event token as follows:

$$P = \text{softmax}(E \cdot W) \in \mathbb{R}^{N \times 2} \quad (4)$$

where $W \in \mathbb{R}^{D \times 2}$ is a trainable weight matrix. During training time, we therefore have the following loss for predictions:

$$\mathcal{L}_{RCE} = CE(P, Y) \quad (5)$$

where Y is the ground-truth label of each event token e_i being the correct answer.

3.3 Event Type Classifier

We describe the implementation of ETC in this subsection. Inspired by the cloze-style prompt learning paradigm for text classification with pre-trained language models, event type classification can be realized by filling the [MASK] answer using a prompt function.

First, the prompt function wraps the input sentence by inserting pieces of natural language text. For prompt function f_p , as illustrated in Fig. 1, we use “[SENTENCE] This sentence describes a [MASK] event” as a prompt function for our model. Let \mathcal{M} be pre-trained language model (i.e., BERT), and \mathbf{x} be the input sentence. The prediction score of each token v in vocabulary being filled in [MASK] token can be computed as:

$$p_v = \mathcal{M}([\text{MASK}] = v | f_p(x)) \quad (6)$$

After that, the other key of prompt learning is answer engineering. We aim to construct a mapping function from event token space to event type space. In the first tower (RCE), it learns the relation between the input sentence and event tokens. RCE and ETC share the same weights of BERT. Then, we only select tokens in $\mathcal{Y} = \{e_1, e_2, \dots, e_N\}$ and compute the scores of event tokens:

$$p_e = \sigma(p_v | v \in \mathcal{Y}) \quad (7)$$

where $\sigma(\cdot)$ determines which function to transform the scores into the probability of event tokens, such as *softmax*.

Finally, as shown in Fig. 1, we predict all event tokens that score higher than the “ $\langle None \rangle$ ” token as the predicted result. In our example, since both “ $\langle assistance \rangle$ ” and “ $\langle employment \rangle$ ” have higher scores than “ $\langle None \rangle$ ”, we predict *Assistance* and *Employment* as target event types.

In the process of training, we calculate two losses due to the problem of imbalance data distribution. The first loss is defined as:

$$\mathcal{L}_1 = \frac{1}{|T|} \sum_{t \in T} \log \frac{\exp(\mathcal{M}([\text{MASK}] = t | f_p(x)))}{\sum_{t' \in \{t, \langle none \rangle\}} \exp(\mathcal{M}([\text{MASK}] = t' | f_p(x)))} \quad (8)$$

where T is the set of event tokens that score higher than “ $\langle None \rangle$ ” in the sentence. The second loss is defined as follows:

$$\mathcal{L}_2 = \log \frac{\exp(\mathcal{M}([\text{MASK}] = \langle none \rangle | f_p(x)))}{\sum_{t' \in \{\langle none \rangle\} \cup \bar{T}} \exp(\mathcal{M}([\text{MASK}] = t' | f_p(x)))} \quad (9)$$

where \bar{T} is the set of event tokens that score lower than “ $\langle None \rangle$ ” in the sentence. Note that in Eq. 8, we only compare the prediction scores that higher than the “ $\langle None \rangle$ ” event token. The reason is that we aim to improve the score of each event token that is higher than “ $\langle None \rangle$ ”. In Eq. 9, we compare to event tokens that lower than the “ $\langle None \rangle$ ”, which can decrease the score of them. The training loss of ETC is defined as:

$$\mathcal{L}_{ETC} = \frac{1}{M} \sum_{x \in \mathcal{S}} (\mathcal{L}_1 + \mathcal{L}_2) \quad (10)$$

In the training time, the total loss of our model is defined as:

$$\mathcal{L} = \mathcal{L}_{RCE} + \mathcal{L}_{ETC} \quad (11)$$

4 Experiments

In this section, we introduce the experimental datasets, evaluation metrics, implementation details, and experimental results.

4.1 Dataset and Evaluation

To evaluate the potential of **EDPRC** under different size datasets, we conducted our experiments on two benchmark datasets, ACE2005 [23] and MAVEN [24]. Details of statistics are available in Table 1.

- The **ACE2005** is globally recognized as the primary multilingual dataset applied for event extraction. Our use focuses on the English version that includes 599 documents and 33 types of events. We engage two versions in line with prior data split pre-processing: ACE05-E [25] and ACE05-E⁺ [26]. In contrast with ACE05-E, ACE05-E⁺ incorporates roles for pronouns and multi-token event triggers.
- **MAVEN**, constructed from Wikipedia² and FrameNet [27], is a vast event detection dataset encompassing 4,480 documents and 168 different types of events.

For data split and preprocessing, following previous work [24–26], we split 599 documents of ACE2005 into 529/30/40 for train/dev/test set, respectively. Then, we use the same processing that splits 4480 documents of MAVEN into 2913/710/857 for train/dev/test set respectively.

To assess the performance of our event detection model, we employ three commonly used evaluation metrics: precision (P), recall (R), and micro F1-score (F1) [2]. These metrics provide a comprehensive picture of our model’s accuracy and effectiveness.

4.2 Baseline

We compare our method to baselines with trigger-based and trigger-free methods. For trigger-based methods, we compare with: (1)**DMCNN** [2], which utilizes a convolutional neural network (CNN) and a dynamic multi-pooling mechanism to learn sentence-level features; (2) **BiLSTM** [28], which uses bi-directional long short-term memory network (LSTM) to capture the hidden states of triggers and classify them into corresponding event types; (3)**MOGANDED** [29], which proposes multi-order syntactic relations in dependency trees to improve event detection; (4)**BERT** [10], fine-tuning BERT on the ED task via a sequence labeling manner; (5)**DMBERT** [4], which adopts BERT as backbone and utilizes a

² <https://www.wikipedia.org/>.

Table 1. Dataset statistics of ACE05-E, ACE05-E⁺ and MAVEN.

Dataset	Split	#Sentences	#Events	#Documents
ACE05-E	Train	17,172	4,202	529
	Dev	923	450	30
	Test	832	403	40
ACE05-E⁺	Train	19216	4419	529
	Dev	901	468	30
	Test	676	424	40
MAVEN	Train	32431	73496	2913
	Dev	8042	17726	710
	Test	9400	20389	857

dynamic multi-pooling mechanism to aggregate textual features. For trigger-free methods, we compare with: (6) **TBNNAM** [6], the first work on detecting events without triggers, which uses LSTM and attention mechanisms to detect events; (7) **TEXT2EVENT** [30], proposing a sequence-to-sequence model and extracting events from the text in an end-to-end manner; (8) **DEGREE** [31], formulating event detection as a conditional generation problem and extracting final predictions from the generated sentence with a deterministic algorithm.

We re-implemented some trigger-based baselines for comparison, including DMCNN, BiLSTM, MOGANDED, BERT and DMBERT. The other baseline results are from the original paper.

4.3 Implementation Details

We utilize the Transformers toolkit [32] and PyTorch to implement our proposed model. Specifically, we employ the *bert-base-uncased*³ model as the backbone and optimize it with AdamW optimizer, setting the learning rate to 2e-5, maximum gradient norm to 1.0, and weight decay to 5e-5. We limit the maximum sequence length to 128 for ACE2005 and 256 for MAVEN, and apply a dropout rate of 0.3. Our model is trained on a single Nvidia RTX 3090 GPU for 10 epochs, selecting the checkpoint with the highest validation performance on the development set. Our code is publicly available at https://github.com/rickltd/event_detection.

4.4 Main Results

Table 2 reports main results. Compared with trigger-free methods, we can find out that our method achieves a much better performance than other trigger-free baselines (TBNNAM, TEXT2EVENT and DEGREE). Obviously, ED_PRC can achieve improvements of 0.4% (73.3% v.s. 73.7%) F1 score of the best trigger-free baseline (DEGREE) in ACE05-E, and 2.1% (71.8% v.s. 73.9%) F1 score of

³ <https://huggingface.co/bert-base-uncased>.

Table 2. Event detection results on both trigger-based and trigger-free methods of the ACE2005 corpora. “-” means not reported in original paper. * indicates results cited from the original paper.

Category	Models	ACE05-E			ACE05-E ⁺		
		P	R	F-1	P	R	F-1
Trigger-based	DMCNN	74.3	66.8	70.3	67.0	73.5	70.1
	BiLSTM	73.6	72.3	72.9	73.5	71.3	72.4
	MOGANED	74.6	71.1	72.8	74.2	72.2	73.2
	BERT	72.5	74.2	73.3	75.2	72.4	73.8
	DMBERT	76.4	71.9	74.1	74.9	73.5	74.2
Trigger-free	TBNNAM*	76.2	64.5	69.9	-	-	-
	TEXT2EVENT*	69.6	74.4	71.9	71.2	72.5	71.8
	DEGREE*	-	-	73.3	-	-	70.9
	ED_PRC (Ours)	76.1	71.5	73.7	74.6	73.2	73.9

TEXT2EVENT in ACE05-E⁺. It proves the overall superiority and effectiveness of our model in the absence of triggers. Compared to trigger-based methods, despite the absence of trigger annotations, ED_PRC can achieve competitive results with other trigger-based baselines, which is only 0.4% (73.7% vs. 74.1%) in ACE05-E and 0.3% (73.9% vs. 74.2%) in ACE05-E⁺ less than the best trigger-based baseline (DMBERT). The result shows that prompt-based method can greatly utilize pre-trained language models to adapt ED task and our MRC module is capable of learning relations between the input text and the target event tokens under low trigger clues scenario.

To further evaluate the effectiveness of our model on large-scale corpora, we show the result of MAVEN on various trigger-based baselines and our model in Table 3. We can see that our model also can achieve competitive performance on various trigger-based baselines, reaching 69.1% F1 score. Compared with CNN-based (DMCNN), RNN-based (BiLSTM) and GNN-based (MOGANED) method, BERT-based methods (BERT, DMBERT and ED_PRC) can outperform high improvements, which indicates pre-trained language models can greatly capture contextual representation of input text. However, ED_PRC can achieve only improvements of 0.1% (67.2% v.s. 67.3%) F1 score on BERT and is 0.8% (67.3% v.s. 68.1%) less than DMBERT. This can be attributed to more triggers and events on MAVEN than that on ACE2005. We conjecture that trigger-based event detection models can greatly outperform trigger-free models when sufficient event information is available. All in all, our ED_PRC is proven competitive in both ACE2005 dataset and MAVEN dataset.

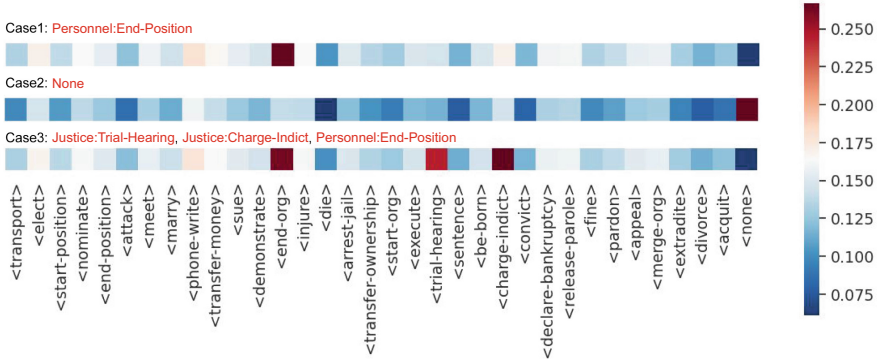


Fig. 2. The ACE2005 examples visualization of attention weight in event tokens. We show three cases, the first with only one event, the second with no events and the third with multiple events.

5 Analysis

In this section, we demonstrate further analysis and give an insight into the effectiveness of our method.

5.1 Effective of Reading Comprehension Encoder

Figure 2 shows a few examples with different target event types and their attention weight visualizations learned by the reading comprehension encoder. In the first case, the target event type is “Personnel:End-Position” and our reading comprehension encoder successfully captures this feature by giving “<end – org>” a high attention score. In addition, in the second case, it is a negative sample that no event happened in this sentence and our reading comprehension encoder can correctly give a high attention score for “<none>” and give low attention scores for other event tokens. Moreover, three events occur in the third case, “Justice:Trial-Hearing”, “Justice:Charge-Indict” and “Personnel:End-Position”, respectively. Our approach can also give high attention scores to “<trial – hearing>”, “<charge – indict>” and “<end – org>”. We argue that, although triggers are absent, our model can learn the relations between input text and event tokens and assign the ground-truth event tokens with high attention scores.

5.2 Effective of Different Prompts

Generally, as the key factor in prompt learning, the prompt can be divided into two categories: hard prompt and soft prompt. The hard prompt is also called a discrete template, which inserts tokens into the original input sentence. Soft prompt is also called continuous template, which is a learnable prompt that does not need any textual templates. To further analyze the influence of prompts, we

Table 3. Event detection results on MAVEN corpus.

Models	P	R	F-1
DMCNN	66.5	58.4	62.2
BiLSTM	64.7	68.2	62.4
MOGANED	65.9	65.1	65.5
BERT	64.3	70.5	67.2
DMBERT	68.9	67.4	68.1
ED_RRC (Ours)	66.0	68.7	67.3

Table 4. Results on ACE2005 datasets with different prompts.

Models	P	R	F-1
Prompt_1	74.2	72.9	73.5
Prompt_2	75.6	71.4	73.4
Prompt_3	74.7	71.2	72.9
Prompt_4	74.1	73.5	73.8
Soft	73.5	72.7	73.1

design four different textual templates (hard prompt) to predict event types: (1) What happened? [SENTENCE] This sentence describes a [MASK] event; (2) [SENTENCE] What event does the previous sentence describe? It was a [MASK] event; (3) [SENTENCE] It was [MASK]; (4) A [MASK] event: [SENTENCE]. For soft prompt, we insert four trainable tokens into the original sentence, such as “[TOKEN] [TOKEN] [SENTENCE] [TOKEN] [TOKEN] [MASK]”. The results of our method on ACE2005 are shown in Table 4.

Prompt_1 and Prompt_2 perform similarly, and both of them work better than Prompt_3. The reason for this may be that Prompt_3 provides less information and less topic-specific. And both Prompt_1 and Prompt_2 add a common phrase “sentence describe” and a question to prompt the model to focus on the previous sentence. Unlike previous prompts, Prompt_4 puts [MASK] at the beginning of a sentence, and the result indicates that it might be slightly better to put the [MASK] at the end of the sentence. Compared with hard prompt, soft prompt eliminate the need for manual human design and construct trainable tokens that be optimized during training time. The result of soft prompt achieve performance that was fairly close to the hard prompt.

6 Conclusion

In this paper, we transform sentence-level event detection to a two-tower model via prompt learning and machine reading comprehension, which can detect events without trigger words. By using machine reading comprehension framework to formulate a reading comprehension encoder, we can learn the relation between input text and event tokens. Besides, we utilize prompt-based learning methods to construct an event type classifier and final predictions are based on two towers. To make effective use of prompts, we design four manual hard prompts and compare with soft prompt. Experiments and analyses show that ED_PRC can even achieves competitive performance compared to mainstream approaches using annotated triggers. In the future, we are interested in exploring more event detection methods without triggers by using prompt learning or other techniques.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China (62006044, 62172110). Additionally, support was also partly provided by the Natural Science Foundation of Guangdong Province (2022A1515010130), and the Programme of Science and Technology of Guangdong Province (2021A0505110004) contributed in part to this work.

References

1. Li, Q., et al.: A survey on deep learning event extraction: approaches and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **PP**, 1–21 (2022)
2. Chen, Y., Xu, L., Liu, K., Zeng, D., Zhao, J.: Event extraction via dynamic multi-pooling convolutional neural networks. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pp. 167–176 (2015)
3. Sha, L., Qian, F., Chang, B., Sui, Z.: Jointly extracting event triggers and arguments by dependency-bridge RNN and tensor-based argument interaction. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32 (2018)
4. Wang, X., Han, X., Liu, Z., Sun, M., Li, P.: Adversarial training for weakly supervised event detection. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 998–1008 (2019)
5. Lai, V.D., Nguyen, T.H., Deroncourt, F.: Extensively matching for few-shot learning event detection. In: *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*, pp. 38–45 (2020)
6. Liu, S., Li, Y., Zhang, F., Yang, T., Zhou, X.: Event detection without triggers. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (2019)
7. Zhang, Z., Kong, X., Liu, Z., Ma, X., Hovy, E.: A two-step approach for implicit event argument detection. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 7479–7485 (2020)
8. Li, X., Feng, J., Meng, Y., Han, Q., Wu, F., Li, J.: A unified MRC framework for named entity recognition. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 5849–5859 (2020)
9. 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 (EACL)*, pp. 255–269 (2021)
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: 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 (NAACL-HLT)*, pp. 4171–4186 (2019)
11. Vaswani, A., et al.: Attention is all you need. In: *Advances in Neural Information Processing Systems* 30 (2017)
12. Ahn, D.: The stages of event extraction. In: *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pp. 1–8 (2006)
13. Cui, S., Yu, B., Liu, T., Zhang, Z., Wang, X., Shi, J.: Edge-enhanced graph convolution networks for event detection with syntactic relation. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 2329–2339 (2020)

14. Yang, S., Feng, D., Qiao, L., Kan, Z., Li, D.: Exploring pre-trained language models for event extraction and generation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 5284–5294 (2019)
15. Wei, Y., et al.: DESED: Dialogue-based explanation for sentence-level event detection. In: Proceedings of the 29th International Conference on Computational Linguistics (COLING), pp. 2483–2493 (2022)
16. Seo, M., Kembhavi, A., Farhadi, A., Hajishirzi, H.: Bidirectional attention flow for machine comprehension. arXiv preprint [arXiv:1611.01603](https://arxiv.org/abs/1611.01603) (2016)
17. Shen, Y., Huang, P.S., Gao, J., Chen, W.: ReasonNet: learning to stop reading in machine comprehension. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1047–1055 (2017)
18. Liu, J., Chen, Y., Liu, K., Bi, W., Liu, X.: Event extraction as machine reading comprehension. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1641–1651 (2020)
19. Zhao, J., Yang, H.: Trigger-free event detection via derangement reading comprehension. arXiv preprint [arXiv:2208.09659](https://arxiv.org/abs/2208.09659) (2022)
20. Brown, T., et al.: Language models are few-shot learners. *Adv. Neural. Inf. Process. Syst.* **33**, 1877–1901 (2020)
21. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.* **55**(9), 1–35 (2023)
22. Wei, Y., Mo, T., Jiang, Y., Li, W., Zhao, W.: Eliciting knowledge from pretrained language models for prototypical prompt verbalizer. In: *Artificial Neural Networks and Machine Learning - ICANN 2022*, pp. 222–233 (2022)
23. Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S., Weischedel, R.: The automatic content extraction (ACE) program - tasks, data, and evaluation. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)* (2004)
24. Wang, X., et al.: MAVEN: a massive general domain event detection dataset. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1652–1671 (2020)
25. Wadden, D., Wennberg, U., Luan, Y., Hajishirzi, H.: Entity, relation, and event extraction with contextualized span representations. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5784–5789 (2019)
26. Lin, Y., Ji, H., Huang, F., Wu, L.: A joint neural model for information extraction with global features. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 7999–8009 (2020)
27. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley Framenet project. In: *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics* (1998)
28. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
29. Yan, H., Jin, X., Meng, X., Guo, J., Cheng, X.: Event detection with multi-order graph convolution and aggregated attention. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5766–5770 (2019)

30. Lu, Y., et al.: Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP), pp. 2795–2806 (2021)
31. Hsu, I.H., et al.: DEGREE: a data-efficient generation-based event extraction model. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 1890–1908 (2022)
32. Wolf, T., et al.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 38–45 (2020)