

# RECOUPLE EVENT FIELD VIA PROBABILISTIC BIAS FOR EVENT EXTRACTION

Xingyu Bai<sup>1\*</sup>, Taiqiang Wu<sup>1\*</sup>, Han Guo<sup>2</sup>, Zhe Zhao<sup>2</sup>, Xuefeng Yang<sup>2</sup>,  
Jiayi Li<sup>1</sup>, Weijie Liu<sup>2</sup>, Qi Ju<sup>2</sup>, Weigang Guo<sup>2</sup>, Yujiu Yang<sup>1</sup> †

<sup>1</sup>Tsinghua Shenzhen International Graduate School, Tsinghua University, China  
<sup>2</sup>Tencent, China

## ABSTRACT

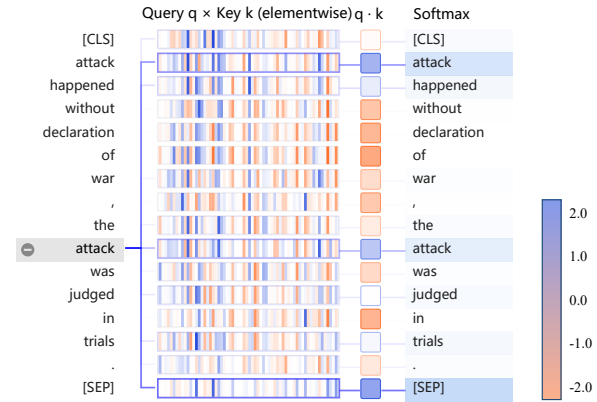
Event Extraction (EE), aiming to identify and classify event triggers and arguments from event mentions, has benefited from pre-trained language models (PLMs). However, existing PLM-based methods ignore the information of trigger/argument fields, which is crucial for understanding event schemas. To this end, we propose a Probabilistic reCoupling model enhanced Event extraction framework (ProCE). Specifically, we first model the syntactic-related event fields as probabilistic biases, to clarify the event fields from ambiguous entanglement. Furthermore, considering multiple occurrences of the same triggers/arguments in EE, we explore probabilistic interaction strategies among multiple fields of the same triggers/arguments, to recouple the corresponding clarified distributions and capture more latent information fields. Experiments on EE datasets demonstrate the effectiveness and generalization of our proposed approach.

**Index Terms**— Event extraction, Pre-trained language model, Probabilistic recoupling

## 1. INTRODUCTION

Event extraction is an essential information extraction (IE) task, aiming to extract event structures from unstructured event mentions. It consists of event detection (ED) and event argument extraction (EAE). For example, in the mention "CNN's Kelly reports on Netanya's attack.", the ED model should identify the event trigger "attack" and classify event type "Conflict", the EAE model should identify the event argument "Netanya" and classify argument role "Place".

Existing EE methods mainly focus on feature engineering. Inspired by the significant performance of PLMs, some prior work [1, 2] utilize general PLMs, such as BERT [3], to construct global dependencies among context words by self-attention. However, PLMs suffer from handling the entanglement of triggers/arguments. As shown in Figure 1, in



**Fig. 1.** Attention score in original self-attention entanglement, connecting line colors are weighted by attention values.

the sentence "Attack happened without declaration of war, the attack was judged in trials.", the first trigger "attack" is more important to the second trigger "attack" than other words when computing self-attention, the information of trigger fields, which is defined as the core auxiliary information scope of triggers, is not highlighted and strengthened enough from the original self-attention entanglement.

To avoid the problem of auxiliary information insufficiency, dependency tree based Graph Convolution Network (GCN) [4, 5, 6] was adopted to capture syntactic relations between triggers and related words. However, this method still has some problems: (1) GCN focuses on the nearest syntactic neighbors [4, 5], over-smoothing [7] in deep layers limits the message passing; (2) GCN has not modeled the event fields and entanglement among the identical event fields.

Another localness-enhanced method, which models central word regions as Gaussian priors [8, 9], could expand the neighbor scopes without changing the model structure.

In this paper, we propose a novel probabilistic recoupling model enhanced EE framework. Specifically, we model and clarify the event fields using probabilistic priors over the corresponding window (i.e., the deviation of syntactic dependency distribution) to the central word (i.e., the position of trigger/argument), and further decouple the original ambiguous self-attention entanglement for the first time.

\* Equal contribution.

† Corresponding author. This work was partly supported by the National Key Research and Development Program of China (No. 2020YFB1708200) and the Shenzhen Key Laboratory of Marine IntelliSense and Computation under Contract ZDSYS20200811142605016.

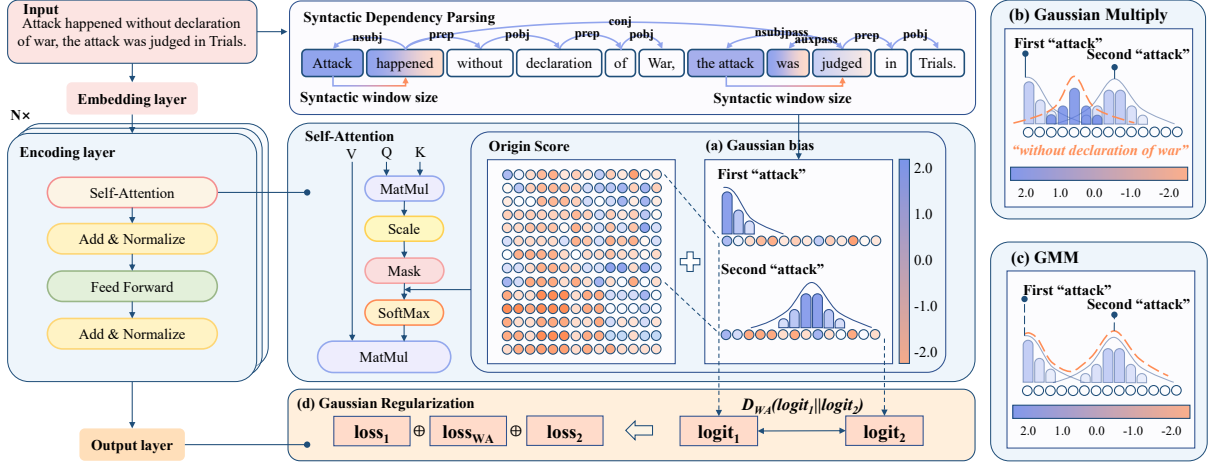


Fig. 2. Illustration of the proposed probabilistic-based event field modeling approach.

According to our statistics on the ACE-2005 dataset, 13.18% (1959/14862) of mentions have the same triggers and arguments occurring multiple times. The same triggers/arguments in the same mentions are identical, modeling the entanglement among multiple occurrences of the same triggers/arguments is beneficial. We explore several probabilistic interaction methods among identical trigger/argument fields to recouple the clarified fields and capture more latent knowledge fields. In addition, we adopt a novel distribution metric loss and regularize the output distributions of the identical triggers/argument fields to be consistent by minimizing the Wasserstein (WA) divergence among the outputs. Experiments on several datasets indicate that ProCE achieves significant performances on overall and few-sample settings.

## 2. METHODOLOGY

The overall ProCE framework consists of three enhancement strategies: (1) Event Field Clarification Modeling, (2) Event Field Recoupling, (3) Event Field Regularization Learning.

**Preprocessing.** To build syntactic-aware event fields, we use the syntactic dependency parser spaCy<sup>1</sup> to parse event mentions into syntactic dependency tree structures, which consist of tokens as nodes and syntactic dependencies as edges.

**Event Field Clarification Modeling.** Self-attention computes attention scores as Figure 1. Specifically, given an event mention  $s = \{w_1, \dots, w_n\}$  contains  $n$  tokens, the hidden state  $H$  is calculated by the transformed queries  $Q \in \mathbb{R}^{n \times d}$ , keys  $K \in \mathbb{R}^{n \times d}$ , and values  $V \in \mathbb{R}^{n \times d}$  as follows:

$$H = \text{Att}(Q, K, V) \quad (1)$$

where  $\text{Att}(\cdot)$  means the dot-product self-attention:

$$\text{Att}(Q, K, V) = \text{softmax}(\text{Score}_{Ori})V \quad (2)$$

$$\text{Score}_{Ori} = \frac{QK^T}{\sqrt{d}} \quad (3)$$

The original calculation regards all words as the same. To efficiently model and clarify trigger/argument field from the original ambiguous entanglement, we suppose that trigger  $t_i$  or argument  $a_i$  represents the central word of event mention  $s$ ; further, we strengthen the importance of tokens in the field of 1-hop syntactic neighbors, which have the most critical syntactic information. Considering the distance contributions of different tokens to central words obey the normal distribution, we model event field via Gaussian prior in this paper.

Specifically, as shown in Figure 2(a), the probabilistic bias  $G \in \mathbb{R}^{n \times n}$ , is defined to clarify and strengthen the original attention score of the event fields, i.e.,  $\text{Score}_{Ori}$  in Equation 2:

$$\text{Att}(Q, K, V) = \text{softmax}(\text{Score}_{Ori} + G)V \quad (4)$$

The element  $G_{i,j} \in (-\infty, 0]$  measures the syntactic distance between context token  $w_j$  in the 1-hop syntactic event field and the central trigger/argument word  $w_i$ , defined as:

$$G_{i,j} = -\frac{(P_j - P_i)^2}{2\sigma_i^2} \quad (5)$$

where  $P_j$  and  $P_i$  are the positions of neighbor and central words,  $\sigma_i$  represents the standard deviation, defined as  $\frac{D_i}{2}$ ,  $D_i$  is the field size of the central word, i.e., the distance of 1-hop syntactic field to specific trigger/argument.

**Event Field Recoupling.** Since the probabilistic bias was introduced independently for central word, it may be beneficial to consider them simultaneously when the same triggers/arguments appear multiple times in the mentions. As shown in Figure 2, the trigger "attack" occurs twice in the mention, and they represent the same event type "war", the interaction between them could provide beneficial field-aware knowledge to each other.

Considering the words simultaneously appear in the interaction of several fields of the same triggers/arguments should be enhanced, we adopt Gaussian multiplication to recouple the clarified fields and obtain the latent distribution:

$$G_{i,j} = -\frac{(P_j - P_i)^2}{2\sigma_i^2} - \frac{(P_j - \mu_{mul})^2}{2\sigma_{mul}^2} \quad (6)$$

<sup>1</sup><https://spacy.io/api/dependencyparser>

### 3. EXPERIMENTS

#### 3.1. Setup

**Datasets.** We evaluate our methods on three generally-used EE datasets, including ACE 2005 English subset [10], which contains 4,090 instances, 33 event types and 35 argument roles; the recently-constructed large-scale OntoEvent dataset [11], which contains 60,546 instances and 100 event types; and FewEvent dataset [12], which contains 70,852 instances and 100 event types. In the ACE 2005 dataset, we evaluate both Event Detection (ED) and Event Argument Classification (EAC) subtasks, while in OntoEvent and FewEvent, we only evaluate ED subtask due to their structures.

**Baselines.** We adopt several official EE baselines, including: (1) CNN-based model DMCNN [13]; (2) the model dependent on syntactic dependency knowledge, such as RNN-based model JRNN [14], GCN-based model JMEE [5], graph-based models DYGIE++ [2], OneIE [15], PathLM [16] and OntoED [11], joint-based model Joint3EE [17]; (3) GAN-based model GAIL [18]; (4) some new forms of EE models, such as QA-based model BERT\_QA [19], generation-based paradigm Text2Event [20]; (5) BERT-based model DMBERT [1].

**General Settings.** We adopt the model structure of BERT, which is with 12 layers, 768 hidden sizes and 12 attention heads. AdamW optimizer is used with the learning rate of  $1 \times 10^{-5}$ . The dimension of token embedding is 768, while the maximum length of the input event mention is 128. The hyperparameters of  $\alpha$  and  $\beta$  are set to 0.5 and 0.5, respectively. We evaluate the performance of EE with Precision (P), Recall (R) and F1 Score (F1). We follow the evaluation protocol of previous EE models, event instances are split into training, validating and testing sets with the ratio of 0.8, 0.1 and 0.1, respectively. We run each method 5 times on all datasets and report the average performances to get stable results.

#### 3.2. Results and Analysis

**Overall Evaluation.** The evaluation results are shown in Tables 1 to 2. We can observe that:

(1) *By modeling trigger/argument syntactic related fields, ProCE could efficiently clarify and recouple event fields from ambiguous entanglement.* ProCE achieves significant improvements compared with DMBERT on all datasets and outperforms all baselines, especially models based on syntactic dependency trees, such as JRNN, JMEE and Joint3EE.

(2) *By combining different probabilistic enhancement strategies, ProCE could achieve different extents of improvements and utilize latent knowledge which benefits EE.* The general ablation study indicates that the recoupling interaction among probabilistic distributions could promote message passing, so the improvements of Gau Fusion, which combines Gau Mul and Gau GMM, are more significant than Gau. Gau Regularization could further improve the performance by strengthening the model’s generalization.

$$\mu_{mul} = \sigma_{mul}^2 \sum_{k=1}^N \frac{\mu_k}{\sigma_k^2}, \quad \frac{1}{\sigma_{mul}^2} = \sum_{k=1}^N \frac{1}{\sigma_k^2} \quad (7)$$

where  $\mu_k$  and  $\sigma_k$  denote the positions and syntactic field sizes of the same triggers/arguments,  $\mu_{mul}$  and  $\sigma_{mul}$  denote the latent probabilistic bias.

To combine several probabilistic distributions of the same triggers/arguments, we adopt GMM strategy to promote the recoupling of co-occurrence syntactic event fields, defined as:

$$G_{i,j} = -\alpha \frac{(P_j - P_i)^2}{2\sigma_i^2} - \sum_{k=1}^N \beta_k \frac{(P_j - P_k)^2}{2\sigma_k^2} \quad (8)$$

where  $\alpha$  and  $\beta_k$  denote the weights of original and other event fields of the same triggers/arguments. Considering exponential operation in *softmax*, we further define GMM as:

$$G_{i,j} = \log\left(\frac{\alpha}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(P_j - P_i)^2}{2\sigma_i^2}\right) + \sum_{k=1}^N \frac{\beta_k}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(P_j - P_k)^2}{2\sigma_k^2}\right)\right) \quad (9)$$

In this way, message passing among multiple distributions could be realized more efficiently.

**Event Field Regularization Learning.** Considering most of the same triggers/arguments occurring multiple times in the same event mentions are identical, we hope the information learned from the identical event fields could be similar. Specifically, the corresponding distributions of predictions  $P_k^{G_k}(y||x_i)$  should be more consistent, where  $G_k$  means the  $k_{th}$  probabilistic distribution of the same triggers/arguments.

Since the output predictions  $P_j^{G_j}(y||x_i)$  and  $P_k^{G_k}(y||x_i)$  of the same triggers/arguments specific fields enhanced knowledge are different, where  $G_j$  and  $G_k$  means the  $j_{th}$  and  $k_{th}$  probabilistic distribution of the same triggers/arguments, the predictions are different for the same input event mention  $x_i$ . To alleviate the inconsistency problem, we adopt a distribution divergence metric, i.e. WA Distance<sup>2</sup>, to regularize the corresponding prediction distributions:

$$L_{WA}^i = \frac{\sum_{j=1}^{N-1} \sum_{k=j+1}^N WA(P_j^{G_j}(y||x_i), P_k^{G_k}(y||x_i))}{N(N-1)/2} \quad (10)$$

With the negative log-likelihood learning objective  $L_{NLL}^i$  of probabilistic enhanced predictions for given data  $(x_i, y_i)$ :

$$L_{NLL}^i = -\frac{\sum_{j=1}^N \log(P_j^{G_j}(y_i||x_i))}{N} \quad (11)$$

Based on these training objectives, we define the final training objective as to minimize  $L_{total}^i$  for the input event mention  $x_i$ :

$$L_{total}^i = L_{NLL}^i + L_{WA}^i \quad (12)$$

In this way, our method further regularizes the model output distributions and enhances model’s generalization. Besides, our method is compatible with the original BERT in model parameters and could be applied conveniently.

<sup>2</sup>where WA Distance could solve the asymmetry problem in Kullback-Leibler (KL) divergence, avoid the bidirectional calculation between two output distributions.

Model	ED			EAC		
	P	R	F1	P	R	F1
DMCNN	75.60	63.60	69.10	<b>62.20</b>	46.90	53.50
JRNN	66.00	73.00	69.30	54.20	56.70	55.40
Joint3EE	68.00	71.80	69.80	52.10	52.10	52.10
DYGIE++	-	-	69.70	-	-	48.80
GAIL	74.80	69.40	72.00	61.60	45.70	52.40
OneIE	-	-	74.70	-	-	56.80
PathLM	-	-	73.40	-	-	56.60
BERT_QA	71.12	73.70	72.39	56.77	50.24	53.31
Text2Event	69.60	74.40	71.90	52.50	55.20	53.80
DMBERT	71.60	72.30	70.87	53.14	54.24	52.76
+Gau	73.43	74.23	72.98	53.82	54.45	53.22
+Fusion	74.22	76.24	74.70	56.35	55.12	54.27
+Regularization	<b>77.16</b>	<b>77.96</b>	<b>76.80</b>	57.43	<b>58.37</b>	<b>57.03</b>

Table 1. Evaluation of EE with various models on ACE2005.

Model	ED		
	P	R	F1
DMCNN	62.51	62.35	63.72
JRNN	63.73	63.54	66.95
JMEE	52.02	53.80	68.07
AD-DMBERT	67.35	73.46	71.89
OneIE	71.94	68.52	71.77
PathLM	73.51	68.74	72.83
OntoED	75.46	70.38	74.92
DMBERT	80.00	78.30	78.40
+Gau	80.45	79.80	79.70
+Fusion	81.18	80.32	80.30
+Regularization	<b>82.30</b>	<b>80.57</b>	<b>80.90</b>

Table 2. Evaluation of ED on OntoEvent.

**Few-sample Evaluation.** Considering that auxiliary information will be more urgent in low-resource EE scenarios, we also study how to influence the ED performance of our method on the FewEvent dataset when trained with different ratios of randomly-sampled training data. We can see that:

(1) *ProCE is especially beneficial for extremely low-resource EE scenarios.* As shown in Table 3, the improvements of our methods compared to the baseline DMBERT are generally more significant when less training data is available. Specifically, in extremely low-resource EE scenarios (training with less than 20% data), Gau Fusion obtains 37.23% F1 with 10% training data, while Gau Regularization obtains 47.13% F1, in comparison to 34.93% in DMBERT.

(2) *ProCE could achieve better performance with less data constantly.* As shown in Table 3, ProCE obtains more advanced performances with less training data continuously. Especially, DMBERT requires 100% training data to almost achieve the best performance, while Gau Fusion only needs 80%. Gau Fusion could even obtain 81.65% F1 with overall data, while Gau Regularization obtains 82.43%, 1.91% higher than 80.52% in DMBERT.

**Detailed Ablation Analysis.** To evaluate the effect of different probabilistic-based enhancement methods, we study the

Model	ED						
	5%	10%	20%	40%	60%	80%	100%
DMBERT	24.37	34.93	56.27	72.47	78.17	79.10	80.52
+Gau	24.93	35.80	56.48	72.77	78.30	79.75	81.30
+Fusion	25.97	37.23	57.25	73.67	79.03	80.23	81.65
+Regularization	<b>32.03</b>	<b>47.13</b>	<b>62.47</b>	<b>75.53</b>	<b>80.13</b>	<b>81.17</b>	<b>82.43</b>

Table 3. Few-sample evaluation with F1-score (%) performance of ED with various models on FewEvent.

Model	ACE ED			ONTO ED			FEW ED		
	P	R	F1	P	R	F1	P	R	F1
DMBERT	71.60	72.30	70.87	80.00	78.30	78.40	81.75	81.83	80.52
+Gau	73.43	74.23	72.98	80.45	79.80	79.70	82.27	82.58	81.30
+Mul	73.39	75.24	73.56	81.16	80.29	80.13	82.48	82.85	81.55
+GMM	73.18	74.78	73.28	81.00	79.98	79.85	82.78	82.62	81.46
+Fusion	74.22	76.24	74.70	81.18	80.32	80.30	82.53	83.00	81.65
+WA	76.52	<b>78.00</b>	76.30	81.90	80.10	80.23	<b>83.57</b>	82.83	81.83
+WA & logits	<b>77.16</b>	77.96	<b>76.80</b>	<b>82.30</b>	<b>80.57</b>	<b>80.90</b>	83.17	<b>83.80</b>	<b>82.43</b>

Table 4. Ablation study of EE on three datasets.

performances of the corresponding modules. The ablation results are shown in Table 4. We can observe that:

(1) *Event Field Recoupling strategy could guide the generation of more syntactic knowledge.* The F1 of Gau is 1.16% higher than DMBERT on the average of all datasets. The Gau Mul slightly surpasses the Gau GMM since Gau Mul generates more unseen information. The fusion of Gau Mul and Gau GMM indicates that promoting message passing among several identical event fields could further benefit the model. The F1 of Gau Fusion is 0.93% higher than Gau on average.

(2) *Event Field Regularization strategy is essential for model’s generalization.* The comparison between Gau WA and Gau WA & logits shows the importance of adjusted probabilistic logit guidance. Since Gau WA only relies on the regularization of probabilistic prediction distributions, the improvements are implicit; Gau WA even results in a 0.07% performance drop by comparison with Gau Fusion on the OntoEvent dataset. Gau WA & logits utilizes all probabilistic enhanced logits to further improve regularization. The F1 of Gau WA & logits is 0.49% higher than Gau WA on the average of all datasets, 3.65% higher than DMBERT on average.

## 4. CONCLUSION AND FUTURE WORK

In this paper, we propose ProCE, a probabilistic recoupling enhanced EE framework, to model the syntactic-related event field knowledge for the first time. To explore the latent entanglement among probabilistic distributions, we construct event field recoupling and regularization methods. Experiments on several datasets demonstrate that our model achieves significant improvements above the previous state-of-the-art models, especially those based on PLMs and syntactic knowledge. Further experiments in few-sample scenarios indicate that our model benefits low-resource EE tasks.

We will expand our methods to other keyword-based IE tasks, such as relation extraction, to improve the generalization ability of our model for future work.

## 5. REFERENCES

- [1] Xiaozhi Wang, Xu Han, Zhiyuan Liu, Maosong Sun, and Peng Li, “Adversarial training for weakly supervised event detection,” in *Proceedings of the NAACL-HLT 2019, Volume 1 (Long and Short Papers)*, 2019, pp. 998–1008.
- [2] David Wadden, Ulme Wennberg, Yi Luan, and Hananeh Hajishirzi, “Entity, relation, and event extraction with contextualized span representations,” in *Proceedings of the EMNLP-IJCNLP 2019*, 2019, pp. 5784–5789.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the NAACL-HLT 2019, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.
- [4] Thien Nguyen and Ralph Grishman, “Graph convolutional networks with argument-aware pooling for event detection,” in *Proceedings of the AAAI 2018, IAAI 2018, EAAI 2018*, 2018, vol. 32.
- [5] Xiao Liu, Zhunchen Luo, and He-Yan Huang, “Jointly multiple events extraction via attention-based graph information aggregation,” in *Proceedings of the EMNLP 2018*, 2018, pp. 1247–1256.
- [6] Taiqiang Wu, Xingyu Bai, Weigang Guo, Weijie Liu, Siheng Li, and Yujiu Yang, “Modeling fine-grained information via knowledge-aware hierarchical graph for zero-shot entity retrieval,” in *Proceedings of the WSDM 2023*, Tat-Seng Chua, Hady W. Lauw, Luo Si, Evimaria Terzi, and Panayiotis Tsaparas, Eds. 2023, pp. 1021–1029, ACM.
- [7] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun, “Graph neural networks: A review of methods and applications,” *AI Open*, vol. 1, pp. 57–81, 2020.
- [8] Baosong Yang, Zhaopeng Tu, Derek F Wong, Fandong Meng, Lidia S Chao, and Tong Zhang, “Modeling locality for self-attention networks,” in *Proceedings of the EMNLP 2018*, 2018, pp. 4449–4458.
- [9] Maosheng Guo, Yu Zhang, and Ting Liu, “Gaussian transformer: a lightweight approach for natural language inference,” in *Proceedings of the AAAI 2019, the IAAI 2019, the EAAI 2019*, 2019, vol. 33, pp. 6489–6496.
- [10] Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda, “Ace 2005 multilingual training corpus,” *Linguistic Data Consortium, Philadelphia*, vol. 57, pp. 45, 2006.
- [11] Shumin Deng, Ningyu Zhang, Luoqiu Li, Chen Hui, Tou Huaixiao, Mosha Chen, Fei Huang, and Huajun Chen, “Ontoed: Low-resource event detection with ontology embedding,” in *Proceedings of the ACL/IJCNLP 2021 (Volume 1: Long Papers)*, 2021, pp. 2828–2839.
- [12] Shumin Deng, Ningyu Zhang, Jiaojian Kang, Yichi Zhang, Wei Zhang, and Huajun Chen, “Meta-learning with dynamic-memory-based prototypical network for few-shot event detection,” in *Proceedings of the WSDM 2020*, 2020, pp. 151–159.
- [13] Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao, “Event extraction via dynamic multi-pooling convolutional neural networks,” in *Proceedings of the ACL/IJCNLP 2015 (Volume 1: Long Papers)*, 2015, pp. 167–176.
- [14] Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman, “Joint event extraction via recurrent neural networks,” in *Proceedings of the NAACL-HLT 2016*, 2016, pp. 300–309.
- [15] Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu, “A joint neural model for information extraction with global features,” in *Proceedings of the ACL 2020*, 2020, pp. 7999–8009.
- [16] Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji, Jonathan May, Nathanael Chambers, and Clare Voss, “Connecting the dots: Event graph schema induction with path language modeling,” in *Proceedings of the EMNLP 2020*, 2020, pp. 684–695.
- [17] Trung Minh Nguyen and Thien Huu Nguyen, “One for all: Neural joint modeling of entities and events,” in *Proceedings of the AAAI 2019, the IAAI 2019, the EAAI 2019*, 2019, vol. 33, pp. 6851–6858.
- [18] Tongtao Zhang, Heng Ji, and Avirup Sil, “Joint entity and event extraction with generative adversarial imitation learning,” *Data Intelligence*, vol. 1, no. 2, pp. 99–120, 2019.
- [19] Xinya Du and Claire Cardie, “Event extraction by answering (almost) natural questions,” in *Proceedings of the EMNLP 2020*, 2020, pp. 671–683.
- [20] Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen, “Text2event: Controllable sequence-to-structure generation for end-to-end event extraction,” in *Proceedings of the ACL/IJCNLP 2021 (Volume 1: Long Papers)*, 2021, pp. 2795–2806.