Hierarchical Event-Relevant Knowledge Injection for Event Argument Extraction

Lexin Ding[1], Shuchao Pang[1,*], Xuxu Ma[1], Bing Li[2,*]
[1] School of Cyber Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[2] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China
{lexinding, pangshuchao, 921129670129}@njust.edu.cn, bing_li@uestc.edu.cn

***Abstract*—Recent Event Argument Extraction (EAE) approaches incorporate auxiliary event-relevant knowledge to improve performance. However, existing methods (1) lack a systematic taxonomy of event-relevant knowledge and (2) exhibit limited knowledge coverage and effectiveness. To address these limitations, we hierarchically decompose event-relevant knowledge into Event Commonsense Knowledge, which is event-type-agnostic, and Inter-Event Correlation Knowledge, which captures transferable relational patterns among related events. A Two-Phase, Parameter-Isolated Training Paradigm inject these knowledge types into distinct parameter subspaces, enabling abstract knowledge acquisition while mitigating catastrophic forgetting. Moreover, a Meta-role Enhanced EAE Prompt and an Event-Correlation Capture Module enable adaptive activation of relevant knowledge conditioned on context, enhancing both the diversity and effectiveness of injected knowledge. Experiments on RAMS, WIKIEVENTS, and MLEE demonstrate consistent improvements over strong baselines, indicating deeper utilization of event-relevant knowledge. Extensive ablation and cross-dataset analyses further confirm the effectiveness and strong scalability of our framework.**

***Index Terms*—Event Argument Extraction; Prompt Learning; Parameter-Efficient Fine-Tuning**

I. Introduction

Event Argument Extraction (EAE) is a core task in information extraction that aims to identify arguments associated with a given event, which benefits many downstream applications such as information retrieval, question answering, and event graph reasoning. As illustrated in Fig. 1, given the event type "Conflict.Attack" triggered by "battle", EAE seeks to extract event arguments along with their corresponding semantic roles, e.g., "police" (Attacker), "rebels" (Target), and "valley" (Place).

Recent advances in EAE have been largely driven by pre-trained language models (PLMs) with prompt-based formulations [1]–[4], which exploit the implicit knowledge encoded in large-scale pretraining corpora. To further enhance performance, a line of research incorporates auxiliary event-relevant knowledge into PLMs, either by retrieving relevant instances as external evidence [5]–[7] or by explicitly modeling dependencies among co-occurring events [8], [9].

Despite notable improvements, existing knowledge-enhanced methods face two fundamental limitations. *First*,

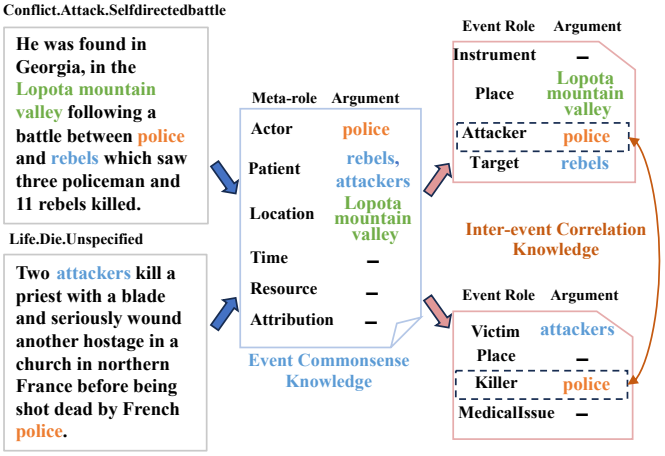

Fig. 1. The illustration of two kinds of auxiliary event-relevant knowledge for EAE.

they lack a *systematic taxonomy of event-relevant knowledge*. Most methods implicitly learn knowledge from instance-level patterns or specific event types, resulting in shallow structural representations and limited generalization across events. Without explicitly disentangling and abstracting different forms of knowledge, such approaches struggle to acquire transferable event semantics beyond observed instances. *Second*, existing methods exhibit *limited knowledge coverage and suboptimal utilization*. Retrieval-based approaches depend heavily on the availability of retrieved examples and are constrained by the limited number. In contrast, dependency modeling methods rely on event co-occurrence signals and thus fail to provide useful guidance for isolated occurring events. Consequently, the injected knowledge covers only a subset of instances and cannot be reliably activated across diverse contexts.

To address these issues, we propose a *Hierarchical Event-Relevant Knowledge Injection Framework via Two-Phase Parameter-Isolated Fine-Tuning* that explicitly decomposes, injects, and activates event-relevant knowledge in a structured manner. Our method hierarchically decompose event-relevant knowledge into: *i*) **Event Commonsense Knowledge**, which is event-type-agnostic and captures the fundamental semantic components shared across events (e.g., *who/what* performs an action, *who/what* is affected, *where* and *when* it occurs); and *ii*)

* Corresponding author.

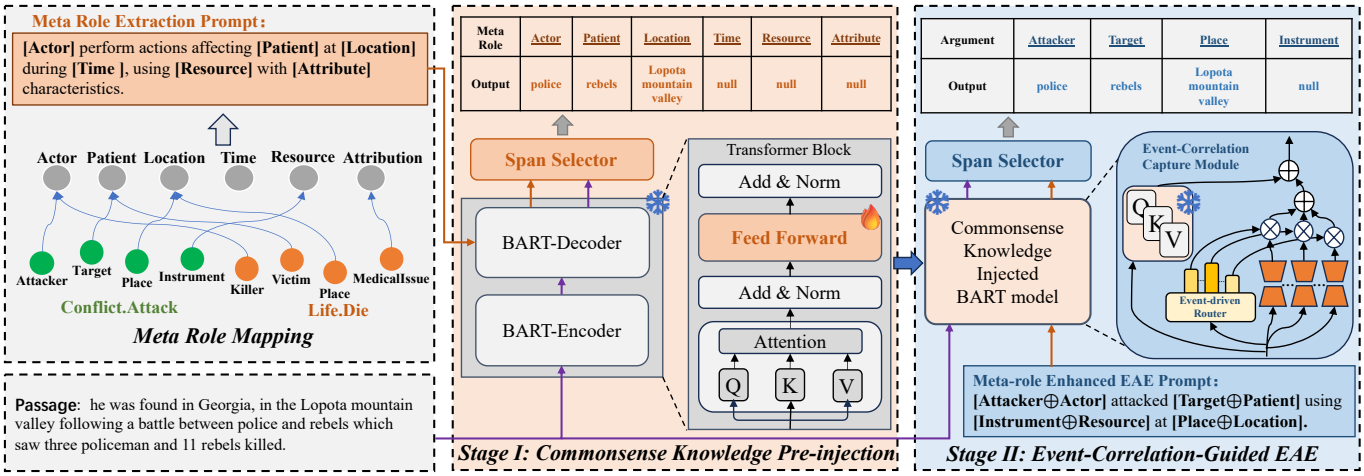

Fig. 2. An overview of our framework.

Inter-Event Correlation Knowledge, which models semantic dependencies and similarities among related event types, as shown in Fig. 1, the roles Attacker and Killer from different event types exhibit strong semantic correspondence, enabling transferable knowledge across events. To model commonsense knowledge, we unify event-specific argument roles into six event-type-agnostic *meta-roles* (i.e., Actor, Patient, Location, Time, Resource, Attribute), as illustrate in Fig. 1, which serve as a shared semantic abstraction for learning universal event structures. To capture correlation knowledge, we introduce an *Event-Correlation Capture Module* composed of multiple LoRA experts and an event-driven dynamic routing mechanism, enabling adaptive and effective knowledge transfer.

Methodologically, building upon pre-trained Encoder-Decoder Transformer (BART [10] or T5 [11]), we design a two-phase parameter-isolated fine-tuning strategy to inject different types of knowledge while mitigating catastrophic forgetting. (1) **Commonsense Knowledge Pre-injection Phase**: We exclusively activate feed-forward network (FFN) parameters as storage modules for commonsense knowledge while freezing all other parameters. Unified prompt templates based on the six meta-roles are constructed as training targets. Owing to the event-type-agnostic nature of commonsense knowledge, this phase is trained on aggregated data from multiple datasets, enabling broad knowledge "pretraining"; (2) **Event-Correlation-Guided EAE Phase**: We design a meta-role-enhanced EAE prompt that explicitly activates the injected commonsense knowledge during EAE. To capture inter-event correlation, we introduce multiple LoRA experts, each specializing in a cluster of semantically related event types. An event-driven router dynamically selecting the most relevant experts for each input instance. Through this hierarchical knowledge decomposition and parameter-isolated injection paradigm, our framework achieves deeper abstraction, broader knowledge coverage, and more effective knowledge utilization for EAE.

Our main contributions are: *i*) a systematic taxonomy of event-relevant knowledge for EAE, distinguishing commonsense and inter-event correlation knowledge; *ii*) a two-phase, parameter-isolated knowledge injection framework that enables structured learning and dynamic activation of event-relevant knowledge while avoiding catastrophic forgetting; *iii*) comprehensive experiments on RAMS, WIKIEVENTS, and MLEE with ablation and multi-dataset analyses confirming the effectiveness and scalability of our approach.

II. Related Work

A. Event Argument Extraction

Recent approaches to Event Argument Extraction (EAE) primarily rely on fine-tuning pre-trained language models with prompt-based formulations to exploit the rich prior knowledge encoded in PLMs. EQAA [4] transformed the EAE task into a question-answering task, extracting event arguments in an end-to-end way. PAIE [1] designs event-specific templates and adopts a slot-filling formulation to guide the model in argument extraction. PAGD [12] further leverages diffusion models to generate context-aware template representations, motivated by their high-quality generation results [13]–[15]. Some methods [7], [16] explore the use of LLMs, yet perform poorly, likely due to instability of LLM-generated content [17]–[19] and the complexity of event schemas [20]. These methods focus solely on exploiting the knowledge encoded in PLMs, while overlooking the rich event-relevant knowledge present in the dataset.

Some methods incorporate event-relevant prior knowledge to help extraction. Retrieval-based methods [5], [6] retrieve relevant examples to enhance generation, however, they superficial capture pattern from instances, which hinders effective knowledge abstraction. Dependency modeling methods [8], [9] extracte multiple events from a sentence simultaneously to capture semantic correlations between events , however, this approach is effective only in scenarios where events co-occur.

B. Parameter-Efficient Fine-Tuning

Optimizing all PLM parameters requires storing a full fine-tuned model for each downstream task, which becomes prohibitively expensive as model sizes increase. Several parameter-efficient fine-tuning (PEFT) methods [21]–[24], such as Adapters, Prefix-Tuning, and LoRA, have been proposed to efficiently adapt PLMs to downstream tasks by

updating only a small subset of parameters. Inspired by these approaches, we combine two PEFT methods to separate commonsense knowledge into FFN layers and correlation knowledge into LoRA experts, enabling effective transfer learning while mitigating catastrophic forgetting.

III. Task Definition

We formulate EAE task as a prompt-based span extraction problem on dataset D. Each training instance is defined as $(X, t, e, R^{(e)}) \in D$, where X is the context, t denotes the trigger word, e denotes the event type, and $R^{(e)}$ denotes the set of event-specific role types, we aim to extract a set of span A. Each $a^{(r)} \in A$ is a segmentation of X and represents an argument about $r^{(e)} \in R^{(e)}$.

IV. Method

A. Model Overview

As shown in Fig. 2, we decompose EAE into two phases to capture distinct types of event-relevant knowledge.

Phase 1: Commonsense Knowledge. Only the feed-forward (FFN) layers are activated while other parameters are frozen. The model learns fundamental event concepts by mapping event-specific roles into six event-type-agnostic meta-roles. This knowledge is stored in FFN parameters.

Phase 2: Correlation Knowledge. We freeze the backbone and introduce multiple LoRA experts with an event-driven router. This enables the model to assign related event types to the same LoRA experts, capturing correlation knowledge while preserving the commonsense.

B. Commonsense Knowledge Pre-injection Phase

1) Meta-Role Mapping: Event commonsense knowledge aims to capture the fundamental structural patterns of events which is event-type agnostic and shared across all event types. Based on event definitions, we identify six fundamental event meta-roles and manually map each event argument role to its corresponding meta-role. Arguments whose roles cannot be mapped to any meta-role are ignored during extraction. Concrete examples are presented in Table I.

2) Meta-Role Extraction Prompt: The overlapping knowledge is dataset-independent, so the prompts used during the overlap knowledge learning phase are identical across all datasets. Special meta-role tokens indicate the positions that the model is expected to fill, as well as the corresponding meta-role types. Given a context C_i, the model is tasked with identifying the correct span to complete the prompt. If several words are categorized as the same meta-role type, they will be concatenated by "and". If there is an empty set of some types, we fill "null" into prompt to replace the special token. The prompt is designed as follows:

> [*Actor*] perform actions affecting [*Patient*] at [*Location*] during [*Time*], using [*Resource*] with [*Attribute*] characteristics.

[•] denotes dynamic tokens which are initialized from embeddings of their original words and fine-tuned during training.

C. Event-Correlation-Guided EAE Phase

We further capture inter-event correlation knowledge to assist EAE, building upon the learned commonsense knowledge. To ignite the commonsense knowledge contained in the frozen FFN layers, we inherit meta-role special tokens from the first learning phase and build prompts according to the target dataset with those special tokens.

1) Meta-Role Enhanced EAE Prompt: For each event type e_i in target dataset, we refer pre-defined prompt from Ma et al. [1] and replace the textual argument roles with a combination of the argument role and its corresponding meta-role to better leverage the commonsense learned in Phase 1. For example, given an event type e: Life.Die.Unspecified, the renovated prompt can be got as:

> [*Attacker* $\oplus$ *Actor*] attacked [*Target* $\oplus$ *Patient*] using [*Instrument* $\oplus$ *Resource*] at [*Place* $\oplus$ *Location*].

$\oplus$ denotes the element-wise addition of embeddings, which integrates meta-roles into event-specific roles to activate commonsense knowledge acquired in Phase 1.

2) Event-Correlation Capture Module : As illustrated in Fig. 2, we employ multiple LoRA experts, each dedicated to capturing different aspects of correlation knowledge. To make efficient use of them, an event-driven router adaptively assigns events to the most suitable LoRA experts. Since many events share similar underlying semantics, the router clusters correlated events to the same LoRA expert, keeping the number of experts small while promoting knowledge sharing across related event types.

At each layer, the router computes correlation scores between the target event and N parallel LoRA experts using the hidden state of the event type, Formally, for each input $h^e_{k-1} \in \mathbb{R}^{d_{in}}$, a router computes gating logits over all experts and derives gating scores as:

$$g = \text{Softmax}(h^e_{k-1} W_g) \in \mathbb{R}^{\mathbb{N}}, W_g \in \mathbb{R}^{d_{in} \times N}, \quad (1)$$

where g denotes the gating scores, $h^e_{k-1} \in \mathbb{R}^{d_{in}}$ is the event-type representation and $W_g \in \mathbb{R}^{d_{in} \times N}$ is router's trainable parameter matrix. The LoRA experts are activated based on correlation scores for forward pass:

$$h = W_0 x + \sum_{i=1}^{N} g_i E_i(x), \quad (2)$$

where $W_0 \in R^{d_{out} \times d_{in}}$ denotes frozen base weights. g_i is the router weight for the i-th LoRA, E_i is the i-th LoRA expert in the selected set:

$$E(x) = \Delta W x = BAx, \quad (3)$$

where the update matrix $\Delta W \in R^{d_{out} \times d_{in}}$ is factorized into two low-rank matrices $A \in R^{r \times d_{in}}$ and $B \in R^{d_{out} \times r}$, with rank $r \ll min(d_{in}, d_{out})$.

D. Model Architecture

We adopt a Transformer-based PLM (e.g., BART [10]) as backbone. As shown in Fig. 2, the model has two components: Role-Specific Selector Generation, and Span Prediction.

TABLE I
SOME EXAMPLES OF MAPPING EVENT-SPECIFIC ROLES TO META-ROLES IN THREE DATASETS.

Dataset	Event Type	Event Argument Role	Meta-Role Type
RAMS	movement.transportartifact.disperseseparate	Artifact	Patient
		Vehicle	Resource
		Origin	Location
		Destination	Location
		Transporter	Actor
	contact.mediastatement.broadcast	Recipient	Patient
		Communicator	Actor
		Place	Location
WikiEvents	contact.requestcommand.meet	Communicator	Actor
		Recipient	Patient
		Topic	Attribute
		Place	Location
	life.die.unspecified	Victim	Patient
		Place	Location
		Killer	Actor
		MedicalIssue	Attribute
MLEE	cell_proliferation	Cell	Actor
	growth	Anatomical Entity	Actor

1) Role-specific Selector Generation: Given a context X and prompt Pt, this module generates a role-specific span selector θ_k for each slot k in the prompt. We adopt BART as the backbone, which consists of an Encoder $\mathcal{M}_{enc}$ and a Decoder $\mathcal{M}_{dec}$.

The context X is encoded by $\mathcal{M}enc$, while the prompt Pt is processed by $\mathcal{M}dec$ with cross-attention to the encoded context:

$$H_X^{(enc)} = \mathcal{M}enc(X), \tag{4}$$

$$H_X = \mathcal{M}dec(H_X^{(enc)}; H_X^{(enc)}), \tag{5}$$

$$H_{pt} = \mathcal{M}dec(Pt; H_X^{(enc)}), \tag{6}$$

where H_X is the event-oriented context representation and Hpt is the prompt representation.

For each role slot k, we mean-pool the corresponding H_{pt} embeddings to obtain a role feature $\psi_k \in \mathbb{R}^h$. Each ψ_k is then projected into start and end selectors:

$$\psi_k^{(start)} = \psi_k \circ w^{(start)}, \ \psi_k^{(end)} = \psi_k \circ w^{(end)}, \tag{7}$$

where $\theta = [w^{(start)}; w^{(end)}] \in \mathbb{R}^{h\times 2}$ are trainable parameters, and $\circ$ denotes element-wise multiplication.

The final span selector for role k is $\theta_k = [\psi_k^{(start)}; \psi_k^{(end)}]$, which is later used in span prediction.

2) Span Prediction: Given the span selector $\theta_k = [\psi_k^{(start)}; \psi_k^{(end)}]$, we query the context representation H_X to identify the most likely argument span (s_k, e_k) for role slot k:

$$\text{logit}_k^{start} = \psi_k^{(start)} H_X \in \mathbb{R}^L, \tag{8}$$

$$\text{logit}_k^{end} = \psi_k^{(end)} H_X \in \mathbb{R}^L, \tag{9}$$

$$\text{score}_k(l, r) = \text{logit}_k^{start}(l) + \text{logit}_k^{end}(r) \tag{10}$$

$$(s_k, e_k) = \arg\max_{(l,r)\in\mathcal{C}} \text{score}_k(l, r), \tag{11}$$

where candidate span set $\mathcal{C} = \{(l, r) \mid (l, r) \in n^2, 0 < r - l < MSL\} \cup \{(0, 0)\}$. MSL denotes the max span length.For a role slot that does not relate to any argument, it is assigned the empty span $(0, 0)$.

For training, we minimize the negative log-likelihood of gold spans:

$$P_k^{start} = \text{Softmax}(\text{logit}_k^{start}), \tag{12}$$

$$P_k^{end} = \text{Softmax}(\text{logit}_k^{end}), \tag{13}$$

$$\mathcal{L} = -\sum_X \sum_k [\log P_k^{start}(s_k) + \log P_k^{end}(e_k)]. \tag{14}$$

where X ranges over all contexts and k over all role slots in the template.

V. EXPERIMENTS

A. Experiments Setup

Dataset and Evaluation Metric. We evaluate on three EAE benchmarks: RAMS [25], WikiEvents (WiE) [26], and MLEE [27]. Following prior work, we report Argument Identification F1 (Arg-I) and Argument Classification F1 (Arg-C). Arg-I counts a prediction as correct if its span matches any gold argument, while Arg-C requires both correct span and role type.

Implementation Details. We adopt BART-large [10] as the backbone. In Phase 1, only FFN layers are trainable while others are frozen. In Phase 2, only LoRA Experts Module are tuned per dataset, we had experimented with different LoRA numbers (4, 6, 8) and ranks (16, 64, 256), the best hyperparameters setting are reported: (RAMS: 6 modules, rank 256), (WiE: 8 modules, rank 16), and (MLEE: 6 modules, rank 64).

Baselines. We compare our method with several EAE methods: (1) prompt-based methods: EEQA [4], PAIE [1], SPEAE [28] and DEGAP [29]; (2)LLMs-based methods: HD-LoA [16] and GEIES [7]; (3) multi-event extraction methods: TabEAE [8] and DEEIA [9]; (4) retrieve-based methods: AHR [6] and PAIE-CMR [30].

TABLE II
ARG-I AND ARG-C F1 ON THREE DATASETS. BEST RESULTS IN **BOLD**, SECOND-BEST UNDERLINED.THE PHASE-2 THREE PARAMETER CONFIGURATIONS CORRESPOND TO THE OPTIMAL HYPERPARAMETERS FOR THREE DATASETS. † DENOTES OUR VARIANT TRAINED WITH THREE DATASETS JOINTLY.

Model	Backbone	Para	RAMS		WiE		MLEE	
			Arg-I	Arg-C	Arg-I	Arg-C	Arg-I	Arg-C
EEQA [4]	BERT-l	340M	48.7	46.7	56.9	54.5	68.4	66.7
PAIE [1]	BART-l	406M	56.8	52.2	70.5	65.3	72.1	70.8
PAIE-Joint†	BART-l	406M	55.9	51.8	70.1	65.2	71.7	69.3
SPEAE [28]	BART-l	406M	58.0	53.3	71.9	66.1	–	–
DEGAP [29]	RoBERTa-l	355M	58.5	54.2	72.2	67.1	74.0	73.4
HD-LoA [16]	GPT-4	–	52.4	44.1	–	–	–	–
GEIES [7]	GPT-4	–	49.2	42.0	–	–	–	–
TabEAE [8]	RoBERTa-l	355M	57.3	52.7	71.4	66.5	72.0	71.3
DEEIA [9]	RoBERTa-l	355M	58.0	53.4	71.8	67.0	75.2	74.3
AHR [6]	T5-l	770M	54.6	48.4	69.6	63.4	–	–
PAIE-CMR [30]	BART-l	406M	**59.3**	54.3	72.8	67.9	–	–
Ours	BART-l	phase-1:201M phase-2: 151M, 12M, 37M	**59.3**	**55.1**	**74.2**	**68.9**	**76.3**	**75.8**

B. Main Results

Table II compares our method with baselines. Since commonsense knowledge is universal across event types, we extend Phase-1 training from single-dataset to multi-dataset settings across RAMS, WiE, and MLEE.

Compared to our backbone PAIE, our method gains +5.6%, +5.5%, +4.2% Arg-C in multi-dataset training. Training with all three datasets in Phase-1 yields the best results on RAMS (+3.3%), WiE (+1.5%) and MLEE (+2.0%). This confirms that our two-phase design successfully learns both commonsense and correlation knowledge, boosting EAE performance.

We attempted to train PAIE on a merged multi-dataset setting, which resulted in lower performance compared to the original PAIE. This suggests that implicitly learning overlapping knowledge from datasets with diverse event structures and event types is difficult. In contrast, our method explicitly captures such overlapping knowledge via a transparent meta-role extraction objective, achieving superior performance.

C. Ablation Study

Table III reports ablations. For component ablation: *i*) Meta-role Enhanced EAE Prompt. Removing meta-role prompt reduces performance, confirming that activating commonsense knowledge during extraction improves accuracy. *ii*) Commonsense Knowledge Injection (CKI). Training only with Phase-2 (FFN+ECCM) , the decline confirm the effectiveness of CKI. The performance still improves over PAIE, showing that correlation clustering benefits knowledge transfer across related events. *iii*) Event-Correlation Capture Module (ECCM). Replacing ECCM with trainable attention layers degrades performance, highlighting the router's role in capturing correlations. To assess the need for distinct parameter modules, we compare two variants: *(i) Full-parameter tuning :* a standard BART model fine-tuned with all parameters activated in two phase training. This joint training underperforms our method, showing that mixing commonsense and correlation knowledge in a shared parameter space causes catastrophic forgetting. *(ii) Phase-reversed:* Phase-1 activates only attention layers, while Phase-2 freezes all parameters and introduces dynamic routing into linear layers. This setup proves less effective, suggesting that FFNs are better suited for storing structured commonsense, whereas correlation knowledge is more naturally captured through attention mechanisms.

TABLE III
ABLATION RESULTS (ARG-C F1) ON THREE DATASETS. Δ INDICATES THE DROP RELATIVE TO THE FULL MODEL.

Variant	RAMS		WiE		MLEE	
	F1	Δ	F1	Δ	F1	Δ
Ours	55.1	–	68.9	–	75.8	–
Phase-reversed	54.7	-0.4	67.8	-1.1	75.0	-0.8
Full-parameter tuning	52.0	-3.1	65.5	-3.4	72.6	-3.2
w/o Enhanced EAE prompt	54.5	-0.6	68.2	-0.7	75.2	-0.6
w/o CKI	53.1	-2.0	66.5	-2.4	74.4	-1.4
w/o ECCM	53.8	-1.3	67.3	-1.6	74.1	-1.7

D. Number of Datasets in Multi-datasets

To examine the impact of training set size on commonsense knowledge injection and test transferability, we varied the number of datasets (0–3) used in Phase-1 training. As shown in Table IV, Arg-C F1 scores consistently improve with more datasets, confirming the universality of our knowledge templates. These results demonstrate that event commonsense knowledge generalizes well across diverse event types and transfers effectively between domains. Importantly, performance improvements persist even when the model is trained exclusively on non-target datasets, demonstrating strong cross-dataset transferability of our approach, enabling performance improvements on target datasets with limited data.

TABLE IV
PERFORMANCE UNDER DIFFERENT NUMBER OF DATASETS IN PHASE-1.

Number	0	1			2			3
	-	RAMS	WiE	MLEE	R&W	R&M	W&M	all
RAMS	52.5	54.0	53.2	52.9	54.8	54.3	53.5	55.1
WiE	65.9	66.7	67.6	66.4	68.6	67.1	67.8	68.9
MLEE	73.2	73.7	73.4	75.0	73.8	75.6	75.4	75.8

E. Hyperparameter Settings for LoRA Experts

Figure 3 illustrates the impact of different LoRA expert hyperparameters. We experimented with varying numbers of

experts (4, 6, 8) and ranks (16, 64, 256). Optimal settings depend on the number of correlated event types and dataset size: RAMS, as the largest dataset with the most event types, benefits from higher ranks, suggesting that larger LoRA ranks are necessary for effective knowledge storage. While WiE, containing many semantically related events, requires more experts to capture diverse inter-event correlations.

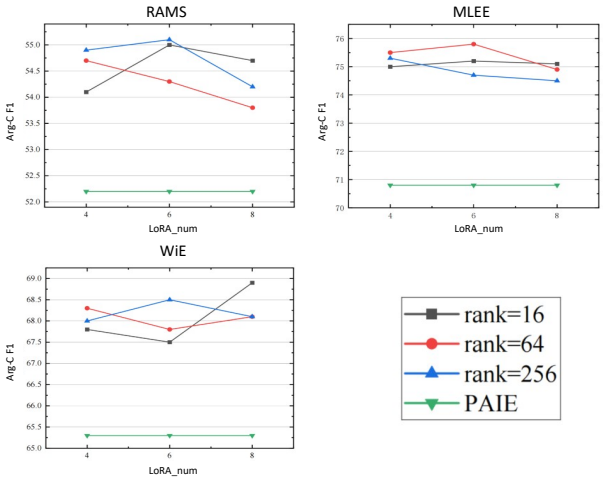

Fig. 3. Impact of different hyperparameter settings on LoRA experts.

VI. Conclusion

We divide event-relevant knowledge for EAE into event commonsense and inter-event correlation, and propose a two-phase parameter-isolated fine-tuning framework. Phase 1 injects commonsense via a meta-role extraction prompt, enhancing generalization. Phase 2 employs an event-driven router with LoRA experts to model transferable event correlations. Experiments on three benchmarks show consistent gains, suggesting the effectiveness and scalability of our approach.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 62476053).

References

[1] Y. Ma, Z. Wang, Y. Cao, M. Li, M. Chen, K. Wang, and J. Shao, "Prompt for extraction? paie: Prompting argument interaction for event argument extraction," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 6759–6774.

[2] I.-H. Hsu, K.-H. Huang, E. Boschee, S. Miller, P. Natarajan, K.-W. Chang, and N. Peng, "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*, 2022, pp. 1890–1908.

[3] X. Gu, L. Lu, X. Zheng, A. Du, Y. Zhou, and S. Pang, "Multimodal robust prompt distillation for 3d point cloud models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 6, 2026, pp. 4320–4328.

[4] X. Du and C. Cardie, "Event extraction by answering (almost) natural questions," *arXiv preprint arXiv:2004.13625*, 2020.

[5] X. Du and H. Ji, "Retrieval-augmented generative question answering for event argument extraction," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022, pp. 4649–4666.

[6] Y. Ren, Y. Cao, P. Guo, F. Fang, W. Ma, and Z. Lin, "Retrieve-and-sample: Document-level event argument extraction via hybrid retrieval augmentation," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 293–306.

[7] Y. Guo, X. Liu, and B. Wu, "Guidelines-enhanced icl with examples selection for document-level event argument extraction," in *2025 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2025, pp. 1–8.

[8] Y. He, J. Hu, and B. Tang, "Revisiting event argument extraction: Can eae models learn better when being aware of event co-occurrences?" in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 12 542–12 556.

[9] W. Liu, L. Zhou, D. Zeng, Y. Xiao, S. Cheng, C. Zhang, G. Lee, M. Zhang, and W. Chen, "Beyond single-event extraction: Towards efficient document-level multi-event argument extraction," in *Findings of the Association for Computational Linguistics ACL 2024*, 2024, pp. 9470–9487.

[10] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7871–7880.

[11] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of machine learning research*, vol. 21, no. 140, pp. 1–67, 2020.

[12] L. Luo and Y. Xu, "Context-aware prompt for generation-based event argument extraction with diffusion models," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 1717–1725.

[13] H. Zhu, S. Pang, Z. Lu, Y. Zhou, and M. Xue, "Gap-diff: protecting jpeg-compressed images from diffusion-based facial customization," in *Network and Distributed System Security Symposium*. Internet Society, 2025.

[14] H. Wang, S. Pang, Z. Lu, Y. Rao, Y. Zhou, and M. Xue, "dp-promise: Differentially private diffusion probabilistic models for image synthesis," in *33rd USENIX Security Symposium (USENIX Security 24)*, 2024, pp. 1063–1080.

[15] S. Pang, Y. Rao, Z. Lu, H. Wang, Y. Zhou, and M. Xue, "Pridm: Effective and universal private data recovery via diffusion models," *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 4, pp. 3259–3274, 2025.

[16] H. Zhou, J. Qian, Z. Feng, L. Hui, Z. Zhu, and K. Mao, "Llms learn task heuristics from demonstrations: A heuristic-driven prompting strategy for document-level event argument extraction," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024, pp. 11 972–11 990.

[17] J. Maynez, S. Narayan, B. Bohnet, and R. McDonald, "On faithfulness and factuality in abstractive summarization," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 1906–1919.

[18] S. Pang, Z. Lu, H. Wang, P. Fu, Y. Zhou, and M. Xue, "Reconstruction of differentially private text sanitization via large language models," in *2025 28th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*. IEEE, 2025, pp. 1–17.

[19] J. Xia, H. Zhu, S. Pang, Z. Lu, B. Li, Y. Zhou, and J. Xue, "One head to rule them all: Amplifying lvlm safety through a single critical attention head," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[20] D. Xu, W. Chen, W. Peng, C. Zhang, T. Xu, X. Zhao, X. Wu, Y. Zheng, Y. Wang, and E. Chen, "Large language models for generative information extraction: A survey," *Frontiers of Computer Science*, vol. 18, no. 6, p. 186357, 2024.

[21] L. Lu, S. Pang, J. Wang, X. Gu, Y. Liu, X. Liu, and Y. Zhou, "Quest: Quantization-conditioned efficient stealthy trojan," *IEEE Transactions on Information Forensics and Security*, 2026.

[22] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for nlp," in *International conference on machine learning*. PMLR, 2019, pp. 2790–2799.

[23] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," *arXiv preprint arXiv:2101.00190*, 2021.

[24] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen *et al.*, "Lora: Low-rank adaptation of large language models." *ICLR*, vol. 1, no. 2, p. 3, 2022.

[25] S. Ebner, P. Xia, R. Culkin, K. Rawlins, and B. Van Durme, "Multi-sentence argument linking," *arXiv preprint arXiv:1911.03766*, 2019.

[26] S. Li, H. Ji, and J. Han, "Document-level event argument extraction by conditional generation," *arXiv preprint arXiv:2104.05919*, 2021.

[27] S. Pyysalo, T. Ohta, M. Miwa, H.-C. Cho, J. Tsujii, and S. Ananiadou, "Event extraction across multiple levels of biological organization," *Bioinformatics*, vol. 28, no. 18, pp. i575–i581, 2012.

[28] C. Nguyen, H. Man, and T. Nguyen, "Contextualized soft prompts for extraction of event arguments," in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 4352–4361.

[29] G. Wang, D. Liu, J.-Y. Nie, Q. Wan, R. Hu, X. Liu, W. Liu, and J. Liu, "Degap: Dual event-guided adaptive prefixes for templated-based event argument extraction with slot querying," in *Proceedings of the 31st International Conference on Computational Linguistics*, 2025, pp. 7598–7613.

[30] W. Liu, E. Zhang, S. Cheng, D. Zeng, L. Zhou, C. Zhang, M. Zhang, and W. Chen, "A compressive memory-based retrieval approach for event argument extraction," in *Proceedings of the 31st International Conference on Computational Linguistics*, 2025, pp. 1278–1293.