ISERA-KGC: Enhancing Knowledge Graph Completion with Interactive Semantic Enhancement and Representation Alignment

Zheping Yu*, Tongxuan Zhang*,†, Guiyun Zhang*
*College of Computer and Information Engineering, Tianjin Normal University, Tianjin, China
Email: yzp@stu.tjnu.edu.cn, txzhang@tjnu.edu.cn, zhangguiyun@tjnu.edu.cn
†Corresponding author

***Abstract*—Knowledge graph completion (KGC) aims to infer missing links by reasoning over triples. Recent work leverages Large Language Models (LLMs) to enhance graph understanding, yet three key challenges remain. First, LLMs often produce ambiguous entity descriptions lacking relational grounding. Second, relation semantics are context-dependent on the conditional entity, a factor frequently overlooked. Third, semantic embeddings from LLMs are misaligned with structural representations, limiting their effectiveness. In this paper, we propose ISERA-KGC (Entity–Relation Interactive Semantic Enhancement and Representation Alignment for Knowledge Graph Completion), a Language Model as a Service framework that leverages frozen LLMs to provide role-sensitive, context-aware semantic descriptions for KGC. Our method introduces two prompting strategies: Role-Aware Relation Prompting (RARP) for generating direction-specific relation semantics, and Relation-Anchored Entity Reasoning (RAER) for grounding ambiguous entities via relational context. Then the Contrastive Fusion and Alignment (CFA) module leverages contrastive learning to align and fuse these semantic and structural embeddings. Experiments on benchmark datasets FB15k-237 and WN18RR show that ISERA-KGC consistently improves mean reciprocal rank (MRR) by up to 2.4 points across multiple KGE models. These results demonstrate the effectiveness of our method in enhancing long-tail reasoning and bridging semantic-structural gaps in knowledge graph completion.**

***Index Terms*—Knowledge Graph Completion, Large Language Models, Prompt Engineering**

I. Introduction

Knowledge graphs (KGs) encode real-world facts as structured triples $\langle head, relation, tail \rangle$ [1], [2], underpinning applications like semantic search [3] and recommendation [4], [5]. However, their inherent incompleteness severely constrains downstream applications, motivating the task of Knowledge Graph Completion (KGC).

Existing KGC methods are mainly embedding-based (e.g., TransE [6], RotatE [7]), which rely on graph structure but struggle with sparse entities, or text-based (e.g., KG-BERT [8], SimKG [9]), which leverage PLMs for semantics but require costly fine-tuning and exhibit domain sensitivity.

To mitigate these limitations, we leverage frozen LLMs via LMaaS [10] to generate contextual descriptions, augmenting KGE models without fine-tuning. However, this poses three challenges: (1) **Role-agnostic semantics**: Existing methods ignore whether an entity acts as a head or tail, yielding generic relation descriptions. (2) **Entity ambiguity**: Without structural context, LLMs fail to ground entities sharing common surface forms (e.g., "Unknown" in Fig. 1), generating noisy descriptions that degrade long-tail representation. (3) **Modality misalignment**: LLM-derived semantic and KGE structural embeddings reside in disjoint latent spaces, hindering integration and introducing noise.

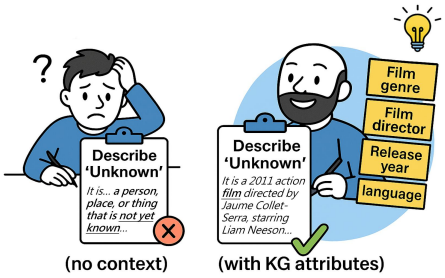

Fig. 1. Entity descriptions without relational context may result in ambiguity (e.g., "Unknown" misread as an adjective). Incorporating KG attributes enables accurate disambiguation.

To address such challenges, we propose the **ISERA-KGC** (**E**ntity–**R**elation **I**nteractive **S**emantic **E**nhancement and **R**epresentation **A**lignment for **K**nowledge **G**raph **C**ompletion) framework, leveraging LLMs to enrich entity semantics and align them with structural representations. Specifically, ISERA-KGC integrates three core components: (1) **Role-Aware Relation Prompting (RARP)** to generate direction-sensitive descriptions; (2) **Relation-Anchored Entity Reasoning (RAER)** to ground LLM generation in local graph attributes; and (3) **Contrastive Fusion and Alignment (CFA)** to project embeddings into a shared space via symmetric contrastive loss, bridging the semantic-structural gap.

In summary, this paper makes the following contributions:

1) We develop a Role-Aware Relation Prompting strategy that constructs prompts based on the conditional entity's role, generating task-specific descriptions to mitigate semantic noise.
2) We introduce a Relation-Anchored Entity Reasoning strategy, leveraging relational context to guide LLMs in identifying correct entities and producing precise, relation-grounded descriptions.
3) We design a Contrastive Fusion and Alignment module projecting semantic and structural embeddings into a unified space via contrastive learning, facilitating fusion.

4) Extensive experiments on FB15k-237 and WN18RR demonstrate that ISERA-KGC achieves state-of-the-art or highly competitive performance.

II. Related Work

A. Traditional KGC Methods

Embedding-based methods represent entities and relations as low-dimensional vectors to capture structural patterns (e.g., TransE [6], RotatE [7]. However, relying solely on structural signals limits their generalization capability on sparse or long-tail entities lacking semantic context. To mitigate this, text-based methods leverage Pretrained Language Models (PLMs) to incorporate semantic information (e.g., KG-BERT [8], SimKGC [9]). Nevertheless, these methods remain limited by their lack of structural awareness and sensitivity to incomplete textual descriptions.

B. LLM-Enhanced KGC

Recent studies leverage LLMs for KGC via generative completion [11], retrieval-augmented reasoning [12], or distillation [13]. However, high inference costs and limited controllability hinder practical deployment, motivating a paradigm shift toward utilizing LLMs as semantic augmenters to complement structural models. Adopting the "LM-as-a-Service" paradigm, methods like MPI-KGC [14] leverage LLMs to generate multi-perspective descriptions. However, these approaches remain constrained by generation noise, entity identification errors, and insensitivity to role-dependent relational semantics.

III. Methodology

As illustrated in Fig. 2, ISERA-KGC aligns LLM-generated semantics with structural embeddings via contrastive learning to score missing entities. To enhance accuracy, our structure-aware prompting strategy integrates entity-level and relation-level contexts through two components.

A. Prompting Strategy

1) Role-Aware Relation Prompting (RARP): A conditional entity's role (head/tail) dictates relation interpretation. Surface-level passive voice transformations (e.g., MPI-KGC [14]) fail on non-predicate relations like *nationality*. RARP overcomes this by generating direction-sensitive prompts. The prompt templates for different entity roles are summarized in Table I. This role-specific guidance improves precision and reduces downstream ambiguity.

2) Relation-Anchored Entity Reasoning (RAER): We transform facts connected to target entity e into semantic attributes, prompting the LLM to reason over specific context for accurate identification. We gather all triples sharing r from training and validation sets:

$$\mathcal{D}_r = \{\, (e', r, e'') \in G_{\text{train}} \cup G_{\text{valid}} \mid e', e'' \in \mathcal{E} \,\} \tag{1}$$

For each e_i, BM25 [15] serves as a lightweight lexical retriever to estimate relevance to r among D_r triples. The top-k relations form an attribute set $\mathcal{A}(e_i, r) = \{r_1, \ldots, r_k\}$. RARP converts these into noun phrases, building a reasoning chain to guide the LLM's inference via context-aware features:

TABLE I
Role-Aware Prompt Templates in RARP.

Conditional Entity Role	Prompt Template
$e = h$ (head)	`Given a triple ([entity_name], [relation], ?), provide an English noun phrase to complete ''[entity_name]'s *''. Output only the noun phrase itself, no extra words or punctuation.`
$e = t$ (tail)	`Given a triple (?, [relation], [entity_name]), provide an English noun phrase to complete ''[entity_name]'s *''. Output only the noun phrase itself, no extra words or punctuation.`

Prompt of Relation-Anchored Entity Reasoning

```
Your task has two stages:
1. Disambiguation:
Use attributes to infer the entity. Then use your
world knowledge to identify the most likely specific
entity that fits the given name and the implied
context. Choose the most prominent entity.
2. Description:
Output a concise and factual description of
the specific entity you have identified. The
description must reflect what is commonly known
about this entity, including aspects implied by the
attributes.
Output Format:
Reasoning:
<step-by-step deduction, from entity type to
specific entity>
Description:
<description of the specific entity>
Entity name: [entity_name]
attributes: [attributes]
```

This CoT prompt grounds generation in structural attributes for precise, relation-specific descriptions.

B. Semantic-Structural Contrastive Fusion and Alignment

To integrate structural and semantic modalities, we design a unified module that executes two key tasks: (i) **contrastive alignment** between LLM-generated and KGE-derived embeddings, and (ii) **representation fusion** to yield a unified scoring vector for downstream prediction.

1) Dual-view Encoding: For $x \in \{h, r, t\}$ in triple (h, r, t), we extract structural embeddings $\mathbf{e}_{\text{stru}}^{(x)}$ and obtain semantic vectors $\mathbf{e}_{\text{sem}}^{(x)}$ by mean-pooling a frozen PLM's outputs on descriptions d_x. Both are projected into a shared space via transformation f_{pro}:

$$\mathbf{z}_{\text{stru}}^{(x)} = f_{\text{pro}}(\mathbf{e}_{\text{stru}}^{(x)}), \mathbf{z}_{\text{sem}}^{(x)} \quad = f_{\text{pro}}(\mathbf{e}_{\text{sem}}^{(x)}), \quad x \in \{h, r, t\} \tag{2}$$

To align both views, we construct a joint matrix $\mathbf{Z}^x \in \mathbb{R}^{2B \times d}$ by row-wise stacking B semantic and B structural vectors, computing similarity:

$$\text{Sim}_{i,j}^x = \frac{\mathbf{Z}_i^x \cdot \mathbf{Z}_j^x}{\tau}, \quad x \in \{h, r, t\} \tag{3}$$

where τ is the temperature hyperparameter. Inspired by CLIP [16], we pair each semantic embedding i with its structural counterpart y_i ($y_i = i + B$ if $i \leq B$, else $i - B$). This bidirectional supervision [17], [18] is optimized via:

$$\mathcal{L}_{\text{CL}}^x = \frac{1}{2B} \sum_{i=0}^{2B-1} -\log \frac{\exp(\text{Sim}_{i,y_i}^x)}{\sum_{j=0}^{2B-1} \exp(\text{Sim}_{i,j}^x)}, \quad x \in \{h, r, t\} \tag{4}$$

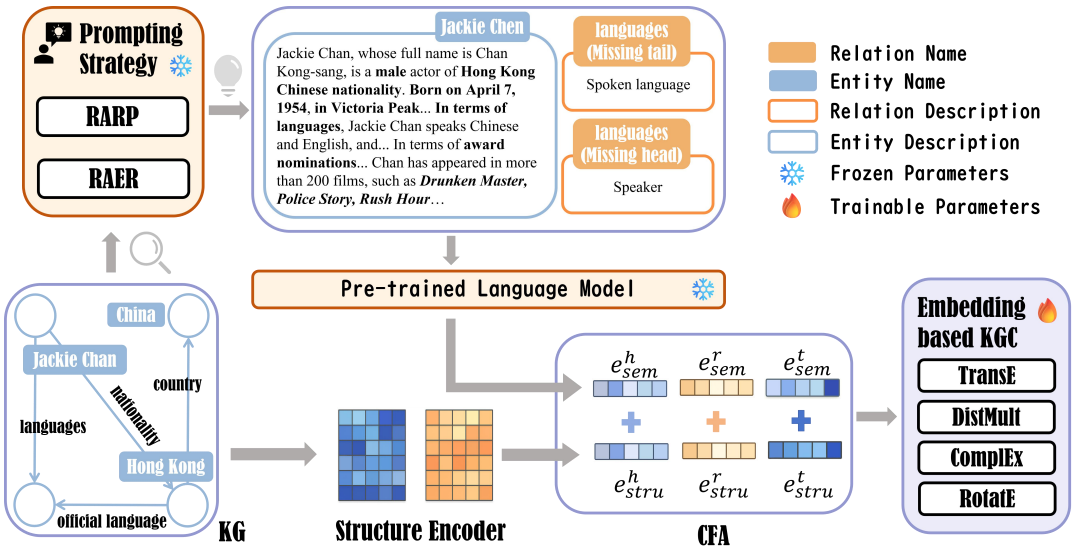

Fig. 2. Overview of ISERA-KGC. RARP and RAER query an LLM for role-aware descriptions, which are encoded by a frozen PLM. The semantic and structural embeddings are then aligned and fused in the contrastive fusion module (CFA). The resulting representations are used by a downstream KGE model for link prediction. Only the KGE components are trainable, all language models remain frozen.

The overall contrastive loss averages these components:

$$\mathcal{L}_{\text{CL}} = \tfrac{1}{3}\left(\mathcal{L}^h_{\text{CL}} + \mathcal{L}^r_{\text{CL}} + \mathcal{L}^t_{\text{CL}}\right) \tag{5}$$

In-batch negative selection avoids global enumeration, while down-weighting ambiguous pairs reduces gradient noise for stable optimization.

Representation Fusion: Post-alignment, we fuse embeddings via linear interpolation:

$$\mathbf{e}^{(x)}_{\text{fus}} = (1-\varepsilon)\,\mathbf{e}^{(x)}_{\text{stru}} + \varepsilon\,\mathbf{e}^{(x)}_{\text{sem}}, \quad x \in \{h, r, t\}. \tag{6}$$

This convex combination (controlled by ε) ensures interpretability and geometric consistency without over-parameterization, unlike complex gating. The fused $\mathbf{e}^{(x)}_{\text{fus}}$ is then processed by the scoring function f_θ.

Semantic-aware Adversarial Negative Sampling: We generate informative negatives via entity corruption. These undergo the identical dual-view pipeline to reside in the same semantic-enhanced space, with weight:

$$w_i \propto \exp\left(\alpha \cdot f_\theta(h'_i, r, t'_i)\right) \tag{7}$$

where temperature α controls sampling sharpness.

Overall Objective: The final objective combines standard KGE loss $\mathcal{L}_{\text{KGC}}$ and alignment loss $\mathcal{L}_{\text{CL}}$ to capture structural and semantic consistency:

$$\mathcal{L} = \mathcal{L}_{\text{KGC}} + \mathcal{L}_{\text{CL}} \tag{8}$$

IV. Experiments

A. Experimental Setup

We describe our experimental setup as follows:

1) Datasets: We evaluate on two benchmarks: WN18RR [19] and FB15k-237 [20]. WN18RR features lexical concepts from WordNet with hierarchical relations, whereas FB15k-237 comprises real-world entities and diverse relations from Freebase [1]. Both datasets are refined versions of WN18 [6] and FB15k [6] with inverse relations removed to mitigate leakage.

TABLE II
Datasets statistics. #Ent. / #Rel.: number of entities / relations. #Train / #Valid / #Test: number of triples.

Dataset	#Ent.	#Rel.	#Train	#Valid	#Test
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134

2) Baselines: In experiments, we adopt several embedding-based KGC models as baselines and apply the proposed model to them. They are: TransE [6], DistMult [26], ComplEX [27], ConvE [19], RotatE [7], HAKE [28]. In addition, we also choose six Text-based models as comparisons, namely KG-BERT [8], KG-S2S [21], LP-BERT [23], KGPT5 [22], SimKGC [9], CSProm-KG [24], and LLMs-based methods namely CD [13], MPIKGC [14], KICGPT [12], COSIGN [25]. Based on ChatGPT, we rerun Zhu's code [29] on the whole dataset and report $\text{ChatGPT}_{\text{zero-shot}}$ and $\text{ChatGPT}_{\text{one-shot}}$ results.

3) Metrics and Implementation Details: We evaluate using MRR and Hits@N ($N \in \{1, 3, 10\}$). We use GPT-4 for descriptions and deberta-v3-base for embeddings on a single RTX 3090. Baselines rely on OpenKE [30] or official codes; grid-searched hyperparameters are detailed in Sec. IV-D.

B. Main Results

We report the KGC results of ISERA-KGC on FB15k-237 and WN18RR in Table III.

1) Overall Performance: ISERA-KGC consistently improves performance across all baselines, with the impact varying by model capacity and graph characteristics. On FB15k-237, structurally simplistic models like TransE see the largest gains (MRR +3.8, Hits@10 +9.5), confirming that contextual semantics effectively compensate for limited structural expressiveness in sparse regions. While high-capacity models (e.g., RotatE) also benefit, their relative improvements are smaller, as their complex geometric rotations already capture most structural patterns. In contrast, gains on WN18RR are

TABLE III

RESULTS ON FB15K-237 AND WN18RR. RESULTS ARE AVERAGED VALUES (OF TEN RUNS FOR FB15K-237 AND WN18RR) OF HEAD/TAIL ENTITY PREDICTIONS. TOP-3 RESULTS FOR EACH METRIC ARE **BOLDED**. ♣ RESULTS REPORTED FROM THE ORIGINAL PAPERS.

		FB15k-237				WN18RR			
Text-based Methods									
Model	PLM	MRR↑	Hits@1↑	Hits@3↑	Hits@10↑	MRR↑	Hits@1↑	Hits@3↑	Hits@10↑
KG-BERT♣ [8]	BERT	—	—	—	42.0	21.6	4.1	30.2	52.4
KG-S2S♣ [21]	T5	33.6	25.7	37.3	49.8	**57.4**	**53.1**	59.5	66.1
KGT5♣ [22]	T5	27.6	21.0	—	41.4	42.6	36.8	—	52.8
LP-BERT ♣ [23]	BERT	31.0	22.3	33.6	49.0	48.2	34.3	56.3	75.2
SimKGC [9]	BERT	32.6	24.1	35.4	49.7	**65.6**	**57.1**	**71.2**	**80.3**
CSProm-KG♣ [24]	BERT	35.2	26.1	38.7	53.6	55.1	50.1	57.0	64.4
LLM-based Methods									
Model		MRR↑	Hits@1↑	Hits@3↑	Hits@10↑	MRR↑	Hits@1↑	Hits@3↑	Hits@10↑
$\text{ChatGPT}_{\text{zero-shot}}$		—	24.8	—	—	—	20.1	—	—
$\text{ChatGPT}_{\text{one-shot}}$		—	29.6	—	—	—	23.7	—	—
GenKGC-CD♣ [13]		—	—	—	—	—	29.3	45.6	53.3
LMKE+MPIKGC-E♣ [14]		30.8	21.9	33.7	48.8	56.4	43.3	**65.5**	**79.5**
KICGPT♣ [12]		**41.2**	**32.7**	**44.8**	55.4	54.9	47.4	58.5	64.1
COSIGN♣ [25]		36.8	31.5	43.4	52.0	**64.1**	**61.0**	65.4	71.4
Embedding-based Methods									
Model	Frame	M_{neg}	MRR↑	Hits@1↑	Hits@3↑	Hits@10↑	MRR↑	Hits@1↑	Hits@3↑	Hits@10↑
TransE	Base [6]	—	27.9	19.8	37.6	44.1	24.3	4.3	44.1	53.2
	ISERA	Org	29.6	20.9	38.9	50.4	28.5	6.0	49.0	55.0
		SANS	31.7	23.2	40.2	53.6	31.0	7.0	51.0	56.5
DistMult	Base [26]	—	28.1	19.9	30.1	44.6	44.4	41.2	47.0	50.4
	ISERA	Org	30.5	22.5	33.8	47.9	45.0	41.7	47.6	51.4
		SANS	29.4	21.0	31.6	45.8	44.8	41.5	47.5	51.0
ComplEx	Base [27]	—	24.7	15.8	27.5	42.8	44.0	41.0	46.0	51.0
	ISERA	Org	26.8	18.0	29.3	44.9	44.3	41.2	46.3	51.4
		SANS	27.8	19.0	30.5	45.8	43.8	40.5	45.8	50.8
ConvE	Base [19]	—	31.2	22.5	34.1	49.7	45.6	41.9	47.0	50.4
	ISERA	Org	34.9	24.0	35.5	50.8	46.8	42.9	48.1	51.3
		SANS	**35.3**	24.2	35.9	51.0	47.2	43.1	48.6	51.7
RotatE	Base [7]	—	33.8	24.1	37.5	53.3	47.6	42.8	49.2	57.1
	ISERA	Org	36.5	30.1	41.6	55.5	54.0	46.0	62.5	74.0
		SANS	**39.3**	**32.4**	**45.2**	**57.0**	56.5	48.0	**66.0**	**75.5**
HAKE	Base [28]	—	34.6	25.0	38.1	54.2	49.7	45.2	51.6	58.2
	ISERA	Org	37.8	28.5	40.0	**56.3**	50.5	45.0	52.0	59.2
		SANS	**41.2**	**32.0**	**45.5**	**57.2**	53.8	50.1	54.8	60.4

TABLE IV

PROMPT ABLATION. "✓" = MODULE ENABLED, "✗" = REPLACED. ABBREVIATIONS: P1 = RARP, P2 = RAER, EN = ENTITYNAME, RN = RELATIONNAME, ME = MPIKGC-E, MR = MPIKGC-R.

Prompt modules						FB15k-237				WN18RR			
P1	P2	EN	RN	ME	MR	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
✓	✓	✗	✗	✗	✗	**39.3**	**32.4**	**45.2**	**57.0**	**56.5**	**48.0**	**66.0**	**75.5**
✓	✗	✓	✗	✗	✗	35.2	26.1	39.4	55.3	50.1	44.5	51.5	56.9
✓	✗	✗	✗	✓	✗	35.4	26.3	39.7	55.6	50.4	44.8	51.9	57.2
✗	✓	✗	✓	✗	✗	34.5	25.6	39.1	54.8	51.3	45.2	52.2	58.0
✗	✓	✗	✗	✗	✓	35.1	26.0	39.6	55.5	52.2	46.3	53.4	59.2
✗	✗	✓	✓	✗	✗	33.8	24.1	37.5	53.3	47.6	42.8	49.2	57.1
✗	✗	✗	✗	✓	✓	35.0	25.8	38.9	54.4	49.4	43.1	50.2	57.0

more modest. This disparity highlights that semantic augmentation yields diminishing returns on structurally dense and semantically unambiguous graphs, whereas it is crucial for heterogeneous, real-world knowledge graphs (like Freebase).

We interpret these results from two perspectives. First, Data Characteristics: FB15k-237 contains diverse relations and ambiguous entities where structure alone is insufficient, making our role-sensitive prompting highly effective. WN18RR's lexical rigor limits this complementary value. Second, Representation Learning: The contrastive objective acts as a semantic regularizer. By aligning the semantic manifold with the structural space, ISERA-KGC enforces consistency that aids generalization to unseen triples. The subsequent fusion serves as a smoothing mechanism, bridging the gap between symbolic structure and linguistic knowledge, which explains the model's robustness on long-tail entities.

TABLE V

ABLATION ON CFA (ROTATE BACKBONE).

Setting	FB15k-237		WN18RR	
	MRR	Hits@10	MRR	Hits@10
Full model	**39.3**	**57.0**	**56.5**	**75.5**
w/o CFA	$29.7_{(\downarrow 9.6)}$	$50.4_{(\downarrow 6.6)}$	$40.2_{(\downarrow 16.3)}$	$51.1_{(\downarrow 24.4)}$

2) Effect of Negative Sampling Strategy: Replacing original sampling with self-adversarial negative sampling (SANS) consistently boosts performance, particularly for translation-based models (TransE, RotatE). These method architectures rely heavily on distinct boundaries, thus benefiting significantly from SANS's focus on harder negatives. In contrast, bilinear models (DistMult, ComplEx) exhibit marginal gains or slight degradation. Their unbounded scoring functions are less sensitive to sampling hardness, which can lead to training instability or diminishing returns, consistent with prior findings [31].

C. Ablation Study

To further elucidate the contribution of each component within the ISERA-KGC framework, we perform ablation studies by systematically removing or modifying key modules. The ablation results, using RotatE as the base model, are presented in Tables IV and V.

1) Prompt Ablation: Table IV shows removing RARP degrades relation-rich FB15k-237 more (MRR $\downarrow$ 4–5), while RAER removal impacts lexical WN18RR heavier ($\downarrow\approx 6$). Removing both causes sharp drops (FB15k-237 $\downarrow$ 4.3, WN18RR $\downarrow$ 7.1). Thus, RARP's relational discrimination and RAER's

TABLE VI
HYPERPARAMETER SENSITIVITY ON ROTATE (FB15K-237). ALL RESULTS ARE AVERAGED OVER 10 RUNS. IMPROVEMENTS OVER THE STRONGEST NON-SANS BASELINE ARE STATISTICALLY SIGNIFICANT UNDER PAIRED T-TEST.

Setting	MRR	Hits@1	Hits@10
τ=0.3, α=0.3	36.9	31.2	54.9
τ=0.3, α=0.5	37.7	31.8	55.8
τ=0.3, α=0.7	38.6	32.4	56.4
τ=0.3, α=1.0	**39.3**	**32.4**	**57.0**
τ=0.5, α=0.3	36.5	30.9	54.6
τ=0.5, α=0.5	37.4	31.5	55.3
τ=0.5, α=0.7	38.5	32.1	56.1
τ=0.5, α=1.0	32.6	39.1	56.7

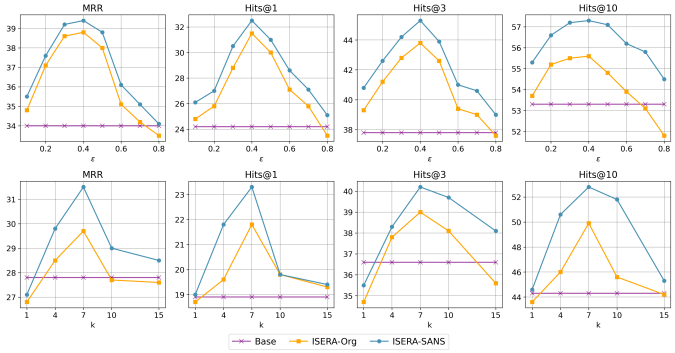

Fig. 3. Analysis of hyperparameter ε and k setting on FB15k-237 (RotatE).

entity grounding provide vital, complementary semantics that structural cues alone cannot substitute.

2) Effect of Contrastive Fusion Alignment (CFA): As shown in Table V, replacing the CFA module with a naive weighted sum results in a severe performance drop (MRR $\downarrow$ 7.4 on FB15k-237, $\downarrow$ 16.3 on WN18RR). This substantial degradation confirms that semantic and structural embeddings naturally reside in distinct latent spaces. Consequently, direct fusion without the alignment enforcement provided by CFA introduces noise, verifying the necessity of contrastive learning for effective modality integration.

D. Hyperparameter Sensitivity

We evaluate four key hyperparameters on FB15k-237 (averaged over 10 seeds): fusion ratio ε, RAER relation count k, adversarial temperature α, and contrastive temperature τ.

Temperatures (α and τ): Table VI shows peak performance at $\alpha = 1.0$ and $\tau = 0.3$ (MRR 39.3, Hits@10 57.0). Sharper adversarial sampling (larger α) helps focus on harder negatives, while a moderate contrastive temperature ($\tau \in [0.3, 0.5]$) stabilizes optimization.

Fusion Ratio (ε) and Relation Count (k): Fig. 3 illustrates inverted-U trends for both parameters, peaking at $\varepsilon \in [0.3, 0.4]$ and $k = 7$ (with 4–10 being optimal). For both, extremes degrade performance: low ε or small k provide insufficient context, while high ε (≥ 0.8) or large k introduce LLM noise that dilutes structural cues [14]. Notably, SANS consistently outperforms the original sampling (Org) across these extremes by prioritizing hard negatives, effectively mitigating semantic noise at high ε and compensating for sparse context at low k, confirming the robustness of our architecture.

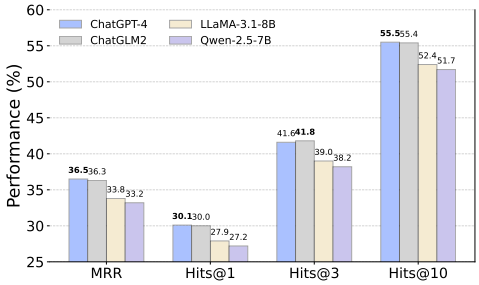

Fig. 4. Comparison of different LLMs.

TABLE VII
EFFECT OF DIFFERENT PLMS ON FB15K-237. THE RESULTS INDICATE THAT THE PROPOSED MODEL REMAINS ROBUST ACROSS DIFFERENT PRETRAINED LANGUAGE MODELS.

PLM	MRR	Hits@1	Hits@3	Hits@10
DeBERTa-v3-base	39.33	32.36	45.18	57.04
EmbeddingGemma	39.41	32.43	45.25	57.06
E5-large	39.28	32.29	45.12	56.99

E. Impact of Base Models (LLMs and PLMs)

We evaluate the framework's sensitivity to the choice of generative LLMs (Fig. 4) and embedding PLMs (Table VII) on FB15k-237.

For LLMs, capacity and instruction tuning are pivotal: GPT-4 achieves the highest MRR (36.5) and Hits@10 (55.5%). Among 7–8B models, LLaMA-3.1-8B outperforms Qwen-2.5-7B, likely because LLaMA's English-centric corpus aligns better with FB15k-237 entities than Qwen's multilingual focus.

For PLMs, substituting `deberta-v3-base` with MTEB leaders [32] (`embeddinggemma` [33] and `e5-large` [34]) alongside RotatE yields minimal performance variance. Ultimately, ISERA-KGC demonstrates consistent improvements and remarkable robustness across diverse pretrained LLM and PLM backbones.

F. Analysis on Long-Tail Entities

Following Wang et al. [35], we evaluate long-tail performance by grouping entities by their logarithmic degrees, assigning triples based on the degree of h or t. Fig. 5 reports Hits@1 and Hits@10 on FB15k-237 across these groups. ISERA-KGC consistently outperforms baselines across almost all groups, achieving significant gains specifically on long-tail entities. This confirms our framework effectively handles relations involving sparse and imbalanced connectivity.

V. CONCLUSION

In this paper, we introduced ISERA-KGC, a framework designed to augment knowledge graph completion by synergizing structural embeddings with role-sensitive, context-aware LLM semantics. By bridging symbolic structure and parametric knowledge via tailored prompting, ISERA-KGC yields consistent gains across standard benchmarks, proving particularly effective for long-tail and ambiguous entities. However, the reliance on frozen LLM services imposes computational and scalability constraints. Accordingly, future work will focus on lightweight or selective prompting strategies to

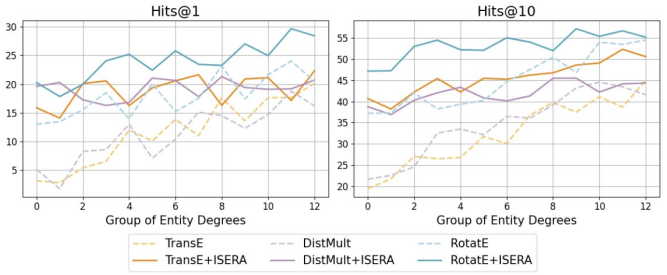

Fig. 5. Predicted performance of RotatE with different LLMs on FB15k-237. Metrics are reported as percentages (%).

optimize cost-effectiveness while preserving performance.

ACKNOWLEDGEMENTS

The research was supported by the Natural Science Foundation of China (Grant No. 62306213) and the Tianjin Normal University Interdisciplinary Research Grant "Fine-grained Student Mental Health Research Based on Large Language Models."

REFERENCES

[1] K. Bollacker, C. Evans *et al.*, "Freebase: a collaboratively created graph database for structuring human knowledge," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008.

[2] F. Mahdisoltani, J. Biega, and F. M. Suchanek, "Yago3: A knowledge base from multilingual wikipedias," in *CIDR*, 2013.

[3] J. Li, H. Peng, and L. Li, "Sublinear smart semantic search based on knowledge graph over encrypted database," *Computers & Security*, vol. 151, p. 104319, 2025.

[4] Z. Rong, L. Yuan, and L. Yang, "Enhanced knowledge graph recommendation algorithm based on multi-level contrastive learning," *Scientific Reports*, vol. 14, no. 1, p. 23051, 2024.

[5] Y. Zhang, K. Chen, X. Bai *et al.*, "Question-guided knowledge graph re-scoring and injection for knowledge graph question answering," *arXiv preprint arXiv:2410.01401*, 2024.

[6] A. Bordes, N. Usunier *et al.*, "Translating embeddings for modeling multi-relational data," *Advances in neural information processing systems*, vol. 26, 2013.

[7] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "Rotate: Knowledge graph embedding by relational rotation in complex space," *arXiv preprint arXiv:1902.10197*, 2019.

[8] L. Yao, C. Mao, and Y. Luo, "Kg-bert: Bert for knowledge graph completion," *arXiv preprint arXiv:1909.03193*, 2019.

[9] L. Wang, W. Zhao, Z. Wei, and J. Liu, "Simkgc: Simple contrastive knowledge graph completion with pre-trained language models," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022.

[10] E. La Malfa, A. Petrov, S. Frieder *et al.*, "Language-models-as-a-service: Overview of a new paradigm and its challenges," *Journal of Artificial Intelligence Research*, vol. 80, pp. 1497–1523, 2024.

[11] L. Yao, J. Peng, C. Mao, and Y. Luo, "Exploring large language models for knowledge graph completion," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[12] Y. Wei, Q. Huang, Y. Zhang, and J. Kwok, "Kicgpt: Large language model with knowledge in context for knowledge graph completion," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 8667–8683.

[13] D. Li, Z. Tan, T. Chen, and H. Liu, "Contextualization distillation from large language model for knowledge graph completion," in *Findings of the Association for Computational Linguistics: EACL 2024*. St. Julian's, Malta: Association for Computational Linguistics, Mar. 2024.

[14] D. Xu, Z. Zhang, Z. Lin *et al.*, "Multi-perspective improvement of knowledge graph completion with large language models," in *LREC-COLING 2024*. 20-25 May, 2024: ELRA Language Resource Association, 2024, pp. 11 956–11 968, cC BY-NC 4.0.

[15] S. E. Robertson and S. Walker, "Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval," in *SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*. Springer, 1994, pp. 232–241.

[16] A. Radford, J. W. Kim, C. Hallacy *et al.*, "Learning transferable visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 139. PMLR, 18–24 Jul 2021.

[17] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.

[18] Z. Huang, X. Jin, C. Lu *et al.*, "Contrastive masked autoencoders are stronger vision learners," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 4, pp. 2506–2517, 2023.

[19] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2d knowledge graph embeddings," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[20] K. Toutanova and D. Chen, "Observed versus latent features for knowledge base and text inference," in *Proceedings of the 3rd workshop on continuous vector space models and their compositionality*, 2015.

[21] C. Chen, Y. Wang, B. Li, and K.-Y. Lam, "Knowledge is flat: A Seq2Seq generative framework for various knowledge graph completion," in *Proceedings of the 29th International Conference on Computational Linguistics*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, Oct. 2022, pp. 4005–4017.

[22] A. Saxena, A. Kochsiek, and R. Gemulla, "Sequence-to-sequence knowledge graph completion and question answering," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 2814–2828.

[23] D. Li, B. Zhu, S. Yang *et al.*, "Multi-task pre-training language model for semantic network completion," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 11, 2023.

[24] C. Chen, Y. Wang, A. Sun, B. Li, and K.-Y. Lam, "Dipping plms sauce: Bridging structure and text for effective knowledge graph completion via conditional soft prompting," *arXiv preprint arXiv:2307.01709*, 2023.

[25] J. Li, H. Yu, X. Luo, and Q. Liu, "Cosign: Contextual facts guided generation for knowledge graph completion," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 2024, pp. 1669–1682.

[26] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *International Conference on Learning Representations (ICLR)*, 2015.

[27] T. Trouillon, J. Welbl *et al.*, "Complex embeddings for simple link prediction," in *International conference on machine learning*. PMLR, 2016, pp. 2071–2080.

[28] Z. Zhang, J. Cai, Y. Zhang, and J. Wang, "Learning hierarchy-aware knowledge graph embeddings for link prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 03, 2020.

[29] Y. Zhu, X. Wang, J. Chen *et al.*, "Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities," *World Wide Web*, vol. 27, no. 5, p. 58, 2024.

[30] X. Han, S. Cao, X. Lv *et al.*, "Openke: An open toolkit for knowledge embedding," in *Proceedings of the 2018 conference on empirical methods in natural language processing: system demonstrations*, 2018.

[31] H. Kamigaito and K. Hayashi, "Comprehensive analysis of negative sampling in knowledge graph representation learning," in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 17–23 Jul 2022, pp. 10 661–10 675.

[32] N. Muennighoff, N. Tazi, L. Magne, and N. Reimers, "Mteb: Massive text embedding benchmark," *arXiv preprint arXiv:2210.07316*, 2022.

[33] H. S. Vera, S. Dua, B. Zhang, D. Salz *et al.*, "Embeddinggemma: Powerful and lightweight text representations," *arXiv preprint arXiv:2509.20354*, 2025.

[34] L. Wang, N. Yang, X. Huang *et al.*, "Text embeddings by weakly-supervised contrastive pre-training," *arXiv preprint arXiv:2212.03533*, 2022.

[35] X. Wang, Q. He, J. Liang, and Y. Xiao, "Language models as knowledge embeddings," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, 7 2022, pp. 2291–2297, main Track.