From Local Perspective to Global Reasoning: A Neuro-Symbolic Framework for Zero-Shot Relation Extraction

Kailun Lyu, Fu Zhang*, Zehan Li, Jingwei Cheng
School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China
{lyukailun, lizehan1999}@163.com, {zhangfu, chengjingwei}@neu.edu.cn

Abstract

Zero-Shot Relation Extraction (ZSRE) aims to predict unseen relations for given entity pairs in sentences. Existing methods typically operate from a local perspective, predicting the relation for each entity pair (given its corresponding sentence) in isolation. Consequently, they often fail to distinguish between unseen, semantically similar relations, particularly when the sentence phrasing is ambiguous. To address this limitation, we propose **G-NSR**, a novel ZSRE framework built upon a **Global Neuro-Symbolic Reasoner** architecture, specifically designed to enable global reasoning across a set of predictions. The key idea is to model the logical relationships among multiple predictions, and perform neuro-symbolic reasoning to ensure logically consistent and more accurate predictions. Specifically, we first introduce Duality Type-Constrained Relation Schemas, which formulate each candidate relation as a pair of complementary positive-negative propositions. These propositions are then synthesized by our designed Neuro-Symbolic Reasoner, which explicitly models their logical interdependencies. By approximating logical rules, the reasoner allows high-confidence predictions to serve as evidence for refining incorrect results, ensuring the final predictions are logically consistent and more accurate. Extensive experiments on widely used datasets demonstrate that our method significantly outperforms existing approaches and establishes new state-of-the-art results across all evaluation settings[1].

1 Introduction

Relation Extraction (RE) is a core natural language processing task that extracts structured triplet facts from unstructured text, aiming to predict a relation between two entities in a sentence. Because many existing methods depend heavily on large annotated corpora (Wadhwa et al., 2023; Zaratiana et al., 2024), interest has grown in Zero-Shot Relation Extraction (ZSRE), where candidate relations at inference are not present in the training data.

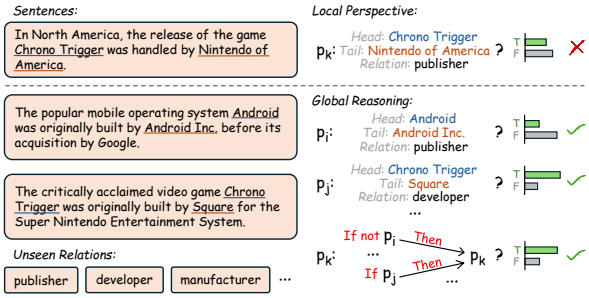

Figure 1: Example contrasting local perspective and global reasoning.

Existing approaches to ZSRE can be broadly categorized into three paradigms. The most common is similarity matching, which aligns sentence and relation embeddings in a shared space (Chen and Li, 2021; Li et al., 2024c; Zhang et al., 2025). A second strategy synthesizes training examples for unseen relations (Chia et al., 2022; Gong and Eldardiry, 2024). More recently, Large Language Models (LLMs) have been used via prompting and in-context learning (Wei et al., 2023; Li et al., 2024a). However, these paradigms are fundamentally limited by their reliance on a **local perspective**, framing the task as a series of independent predictions, where the relation for each entity pair (given its corresponding sentence) is evaluated in isolation. This difficulty increases when the phrasing within the sentence itself is ambiguous, which can lead to confusion between unseen, semantically similar relations. In this case, the model incorrectly predicts the relation between "*Chrono Trigger*" and "*Nintendo of America*" (p_k) as shown in Figure 1.

In contrast, **global reasoning** can alleviate this limitation by treating individual predictions not

*Corresponding author.
[1]Our code is available at https://github.com/LyuKaiLun/G-NSR

Findings of the Association for Computational Linguistics: ACL 2026, pages 29844–29856
July 2-7, 2026

as isolated facts, but as interconnected pieces of evidence that inform one another, thereby improving overall prediction performance. For instance, the predictions on p_i and p_j in Figure 1 serve as strong reference evidence that help clarify the semantics of the "*publisher*" and "*developer*" relations. Aggregating and effectively utilizing evidence from these predictions allows the model to construct a more robust, generalized understanding of sentences and relations. This refined understanding can guide the model to refine its initial errors (e.g., p_k) and infer a more logically consistent and accurate set of predictions.

Motivated by this insight, we introduce a novel framework for ZSRE that is explicitly designed to enable this form of global, interdependent reasoning. Achieving this requires a model that can both capture the rich semantics of sentences and relations and enforce logical consistency across its various predictions. Based on this requirement, inspired by the logical reasoning capabilities of neuro-symbolic systems, we propose **G-NSR**, a novel neuro-symbolic framework for ZSRE that performs **G**lobal reasoning via a ZSRE-specific **N**euro-**S**ymbolic **R**easoner. Specifically, given each entity pair and its corresponding sentence, our approach begins by using **Duality Type-Constrained Relation Schemas** to represent each potential fact (sentence, entity pair, candidate relation) as a verifiable proposition. For each candidate relation, these schemas generate a pair of complementary statements: a positive proposition asserting the relation holds and a negative one asserting its logical opposite. The **Neuro-Symbolic Reasoner** that we designed then operates on this entire collection of propositions. This module is designed to perform global reasoning by utilizing neural networks to represent a series of complex logical rules. Instead of assessing each proposition in isolation, this process allows high-confidence judgments to serve as evidence to refine incorrect results and ensure logically consistent and more accurate predictions, thereby improving ZSRE performance from a global perspective.

Our contributions are summarized as follows:

- We, for the first time, propose a neuro-symbolic ZSRE framework that formulates the ZSRE task as a set of propositions and performs global neuro-symbolic reasoning across them to ensure logically consistent and more accurate predictions.
- We propose Duality Type-Constrained Relation Schemas to construct pairs of logically complementary propositions, providing a structured and robust foundation for the subsequent reasoning process.
- We design a novel Neuro-Symbolic Reasoner that models logical relationships between propositions, enabling it to perform global reasoning for ZSRE.
- Experiments on widely used datasets, Wiki-ZSL and FewRel, show that our framework consistently outperforms prior methods (average **+4.14**) and sets new state-of-the-art results across all evaluation settings. The cost analysis also demonstrates that our framework maintains competitive reasoning efficiency.

2 Related Work

2.1 Zero-Shot Relation Extraction

Most traditional ZSRE methods follow a similarity-matching paradigm, learning a joint embedding space to align sentence representations with relation prototypes (Chen and Li, 2021). Subsequent efforts refine this by improving prototype quality through the aggregation of side information (Li et al., 2024c; Zhang et al., 2025) or by enhancing matching granularity and efficiency with techniques like score decomposition (Zhao et al., 2023) and multi-grained architectures (Li et al., 2024b; Boylan et al., 2025). Another line of work tackles data scarcity by generating synthetic training data for unseen relations (Chia et al., 2022). More recently, LLMs have reframed ZSRE as a prompting or retrieval task, using strategies like meta-tuning for in-context learning (Li et al., 2024a), multi-turn question-answering (Wei et al., 2023), or generation-augmented retrieval (Li et al., 2025).

However, a common characteristic of these approaches is their focus on making predictions for each triplet fact in isolation. They primarily leverage local evidence from a single sentence-relation pair while overlooking cross-fact consistency and the interconnected reasoning evidence across the full set of predictions.

2.2 Neuro-Symbolic Reasoning

Neuro-symbolic reasoning integrates neural networks with symbolic logic to build more interpretable and consistent models. Key techniques include regularizing models with logic-based loss

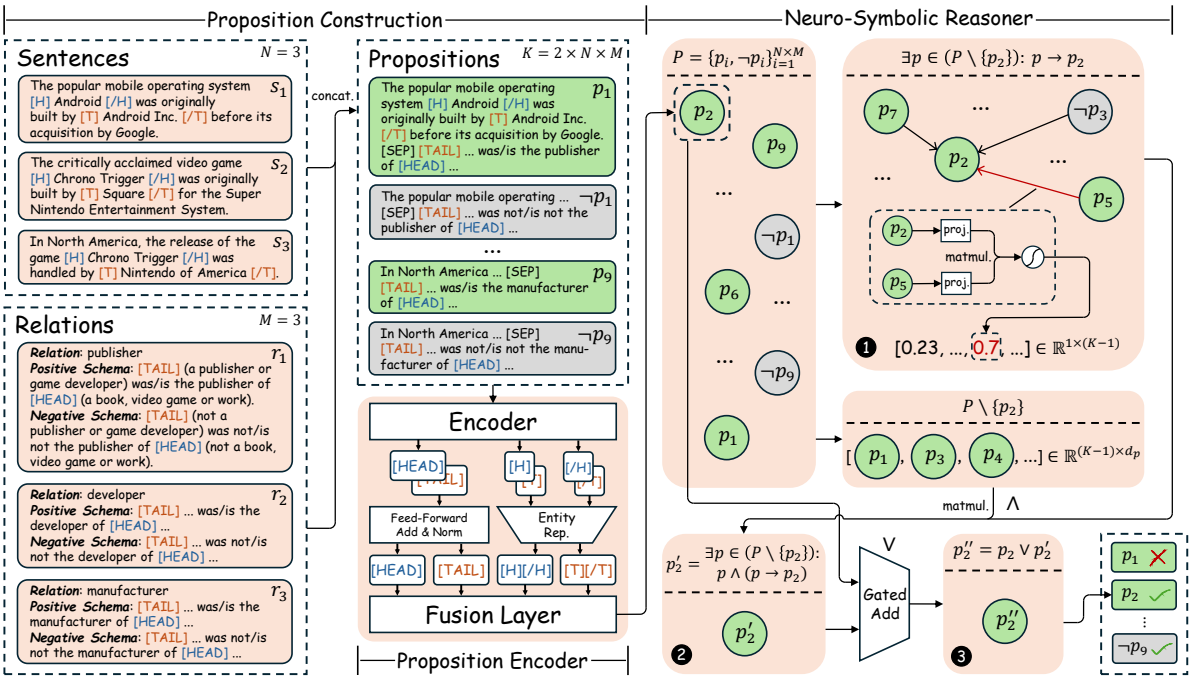

Figure 2: An overview of our proposed framework G-NSR for ZSRE.

functions (Li and Srikumar, 2019; Bosselut et al., 2021), translating problems for external symbolic solvers (Pan et al., 2023; Yasunaga et al., 2024), and designing end-to-end architectures with differentiable modules that approximate logical operations (Rocktäschel and Riedel, 2017; Evans and Grefenstette, 2018; Badreddine et al., 2022).

While existing neuro-symbolic techniques are not designed for ZSRE and cannot be directly adapted to this task, our framework introduces a novel approach specifically tailored to address the local-global challenge of ZSRE mentioned above, ensuring logical consistency across predictions and the ability to generalize to unseen relations.

3 Methodology

3.1 Task Formulation

The task of RE is defined upon a dataset $D = \{(s_i, e_i^{\text{head}}, e_i^{\text{tail}}, r_i)\}_{i=1}^{|D|}$. For each prediction, the input is a sentence s_i containing a head entity e_i^{head} and a tail entity e_i^{tail}, and the goal is to predict the correct relation r_i from a predefined relation set R.

ZSRE applies a specific constraint to this task by splitting the data into a seen set D^s for training and an unseen set D^u for evaluation. The model learns exclusively from D^s and is expected to generalize to D^u. The core challenge of the zero-shot setting is that the set of relations present in the seen data, R^s, is entirely disjoint from the set of relations in the unseen data, R^u ($R^s \cap R^u = \emptyset$).

3.2 Overview

An overview of our proposed framework G-NSR is shown in Figure 2. It includes three main components: (1) **Proposition Construction** utilizes Duality Type-Constrained Relation Schemas to construct a set of paired *positive* (p_i) and *negative* ($\neg p_i$) *propositions* from input sentences and relations. (2) A **Proposition Encoder** then transforms each text-based proposition into a compact vector embedding by fusing key entity and relation features. (3) The **Neuro-Symbolic Reasoner** reasons over the entire set of embeddings through a *three-stage* logical process, modeling logical implications to aggregate global evidence and determine a final, truth-value for each proposition.

3.3 Constructing Propositions Using Duality Type-Constrained Relation Schemas

In our framework, the fundamental unit of reasoning is the *proposition*, defined as a declarative statement constructed for a specific sentence-relation pair that asserts the validity of the relation. To systematically generate these propositions, we first design Duality Type-Constrained Relation Schemas that capture the semantic structure of each relation r. The schema for a relation is formalized as a

	Positive Form (+)	Negative Form (-)
π	was/is the publisher of	was not/is not the publisher of
κ_{HEAD}	a book, video game or work	not a book, video game or work
κ_{TAIL}	a publisher or game developer	not a publisher or game developer

Table 1: An example of the parameterized construction schema for the "*publisher*" relation.

parameterized template:

$$\tau_r(\pi, \kappa_1, \kappa_2) = (\ldots, [E_1](\kappa_1), \ldots, \pi, \ldots, [E_2](\kappa_2), \ldots), \quad (1)$$

where $[E_1]$ and $[E_2]$ are placeholders for the head and tail entities, with their assignment determined by $\{[E_1], [E_2]\} = \{[\text{HEAD}], [\text{TAIL}]\}$. The schema structure is defined by three parameters: π represents a relational predicate, and κ_1, κ_2 impose fine-grained type constraints (typically obtained from ZSRE datasets) on the respective entities.

For each relation r, we define a set of positive parameters $(\pi^+, \kappa^+_{\text{HEAD}}, \kappa^+_{\text{TAIL}})$ and corresponding negative parameters $(\pi^-, \kappa^-_{\text{HEAD}}, \kappa^-_{\text{TAIL}})$. Applying them to the base template τ_r, we generate the positive form τ_r^+ and the negative form τ_r^-:

$$\begin{aligned} \tau_r^+ &= \tau_r(\pi^+, \kappa^+_{\text{HEAD}}, \kappa^+_{\text{TAIL}}), \\ \tau_r^- &= \tau_r(\pi^-, \kappa^-_{\text{HEAD}}, \kappa^-_{\text{TAIL}}). \end{aligned} \quad (2)$$

Here, the negative parameters are defined as the logical opposites of the positive ones (e.g., π^- is the negation of π^+). This ensures that the resulting schemas are structurally symmetric and semantically opposite, providing a dual basis for verification. Table 1 illustrates this process using the "*publisher*" relation. This dual structure allows the model to verify relations from opposing perspectives, ensuring more robust factual judgments.

Then, we mark the head and tail entities in the input sentence s using four boundary tokens ([H], [/H], [T], [/T]), and concatenate with the relation schemas to yield the propositions:

$$\begin{aligned} p &= [s; [\text{SEP}]; \tau_r^+], \\ \neg p &= [s; [\text{SEP}]; \tau_r^-], \end{aligned} \quad (3)$$

where $[\cdot;\cdot]$ denotes concatenation, and the special token [SEP] is a separator.

3.4 Generating Proposition Embeddings

The Proposition Encoder maps a text proposition p_i ($p_i \in \{p, \neg p\}$) into an embedding $\mathbf{v}_{p_i} \in \mathbb{R}^{d_p}$. We first utilize a pre-trained language model BERT-base (Devlin et al., 2019) to extract embeddings for the sentence boundary tokens ($\mathbf{h}_{[\text{H}]}, \ldots, \mathbf{h}_{[\text{/T}]}$) and schema placeholders ($\mathbf{h}_{[\text{HEAD}]}, \mathbf{h}_{[\text{TAIL}]}$). Based on these features, we compute the sentence-level representations $\mathbf{e}_h^{(s)}$ and $\mathbf{e}_t^{(s)}$ for the head and tail entities using an `EntityRep` module:

$$\begin{aligned} \mathbf{e}_h^{(s)} &= \texttt{EntityRep}([\mathbf{h}_{[\text{H}]}; \mathbf{h}_{[\text{/H}]}]), \\ \mathbf{e}_t^{(s)} &= \texttt{EntityRep}([\mathbf{h}_{[\text{T}]}; \mathbf{h}_{[\text{/T}]}]). \end{aligned} \quad (4)$$

Similarly, the schema embeddings are projected into schema-level representations $\mathbf{e}_h^{(\tau)}$ and $\mathbf{e}_t^{(\tau)}$.

Finally, a `Fusion` layer integrates these vectors, yielding the final embedding of proposition p_i:

$$\mathbf{v}_{p_i} = \texttt{Fusion}(\mathbf{e}_h^{(s)}, \mathbf{e}_t^{(s)}, \mathbf{e}_h^{(\tau)}, \mathbf{e}_t^{(\tau)}). \quad (5)$$

Both `EntityRep` and `Fusion` are implemented as Multilayer Perceptrons (MLPs) with residual connections to capture complex feature interactions, and the details are further provided in **Appendix A**.

3.5 Aggregating Global Evidence via Neuro-Symbolic Reasoner

Although a proposition embedding $\mathbf{v}_{p_i}$ can capture the local semantics of a specific sentence-relation pair, it is computed in isolation, overlooking the cross-fact consistency and the interconnected reasoning evidence across the full set of predictions. To bridge this gap, we design a Neuro-Symbolic Reasoner that refines these judgments through global reasoning.

Let P denote the global set of K propositions ($K = 2 \times N \times M$, where N and M denote the number of sentences and relations, respectively), encompassing all positive and negative candidates. The reasoner updates the truth-value logit l_i for each proposition $p_i \in P$ via a *three-stage* logical process as shown in Figure 2.

Logic Implication Induction: We first determine *whether any other proposition p_j in the global set implies the target proposition p_i*. Logically, this corresponds to checking for the existence of a valid implication relationship:

$$\exists p_j \in (P \setminus \{p_i\}) : p_j \rightarrow p_i. \quad (6)$$

To approximate this logical dependency, we model the implication strength $w_{j \rightarrow i}$ by computing the interaction between the projected embeddings:

$$w_{j \rightarrow i} = \sigma\left((\mathbf{W}_1 \mathbf{v}_{p_j} + \mathbf{b}_1)^\top (\mathbf{W}_2 \mathbf{v}_{p_i} + \mathbf{b}_2)\right). \quad (7)$$

This operation yields a weight vector $\mathbf{w}_i \in \mathbb{R}^{1\times(K-1)}$ that quantifies the likelihood of $p_j \rightarrow p_i$ being true.

Evidence Aggregation via Soft Modus Ponens: Next, we *gather global evidence* supporting p_i (denoted as p_i') by applying a soft version of Modus Ponens. Logically, this step seeks to verify whether there exist true premises p_j that imply the conclusion p_i:

$$p_i' = \{p_j \mid \exists p_j \in (P \setminus \{p_i\}) : p_j \wedge (p_j \rightarrow p_i)\}. \tag{8}$$

To approximate this operation, we perform a weighted aggregation of the premise embeddings. Let $\mathbf{V}_{\setminus i} \in \mathbb{R}^{(K-1)\times d_p}$ be the matrix stacking the embeddings of all other propositions. We compute the aggregated evidence vector $\mathbf{v}'_{p_i}$ as:

$$\mathbf{v}'_{p_i} = \mathbf{w}_i \mathbf{V}_{\setminus i}. \tag{9}$$

The matrix multiplication sums up the weighted features, acting as a soft existential quantifier ($\exists$) that combines evidence from all valid premises.

Truth Verification via Soft Disjunction: Finally, we *determine the validity of p_i by synthesizing local and global evidence*. Logically, this corresponds to a disjunction, asserting that p_i holds if it is supported by either its intrinsic local semantics (p_i) or external global evidence (p_i'):

$$p_i'' = p_i \vee p_i'. \tag{10}$$

To approximate this logical operation, we employ an adaptive gating mechanism that dynamically balances the embedding $\mathbf{v}_{p_i}$ with the aggregated evidence $\mathbf{v}'_{p_i}$. We compute a gating coefficient λ_i via a linear projection of the concatenated features:

$$\begin{aligned} \lambda_i &= \sigma(\mathbf{W}_3[\mathbf{v}_{p_i}; \mathbf{v}'_{p_i}] + \mathbf{b}_3), \\ \mathbf{v}''_{p_i} &= \lambda_i \cdot \mathbf{v}_{p_i} + (1-\lambda_i) \cdot \mathbf{v}'_{p_i}. \end{aligned} \tag{11}$$

The resulting vector $\mathbf{v}''_{p_i}$ is then projected through a final linear layer to obtain the scalar logit l_i, which serves as the model's confidence score for the proposition p_i.

3.6 Training and Inference

Training Objective. The model is trained end-to-end by minimizing a composite loss function. For each training instance, we construct a proposition set P consisting of the ground-truth relation and a dynamic selection of negative relations sampled from the same mini-batch.

To enforce logical consistency during reasoning, we introduce an auxiliary Implication Loss $\mathcal{L}_{\text{imp}}$. Let $y_i \in \{0, 1\}$ be the ground-truth label for proposition p_i. We derive the ground-truth label for the implication $y_{j\rightarrow i}$ based on the definition of Material Implication:

$$y_{j\rightarrow i} = \begin{cases} 0 & \text{if } y_j = 1 \text{ and } y_i = 0 \\ 1 & \text{otherwise} \end{cases}. \tag{12}$$

The loss is computed by minimizing the divergence between the predicted weights $w_{j\rightarrow i}$ and the labels:

$$\begin{aligned} \mathcal{L}_{\text{imp}} = -\frac{1}{K(K-1)} \sum_{i\neq j} (y_{j\rightarrow i} \log(w_{j\rightarrow i}) \\ +(1-y_{j\rightarrow i}) \log(1-w_{j\rightarrow i})). \end{aligned} \tag{13}$$

Simultaneously, we optimize the Prediction Loss $\mathcal{L}_{\text{pred}}$ to ensure accurate final judgments. This component directly supervises the final logits l_i against the ground-truth labels y_i:

$$\begin{aligned} \mathcal{L}_{\text{pred}} = -\frac{1}{K} \sum_{i=1}^{K} (y_i \log(\sigma(l_i)) + \\ (1-y_i) \log(1-\sigma(l_i))). \end{aligned} \tag{14}$$

The final training objective is a weighted sum controlled by hyperparameters ζ_1 and ζ_2:

$$\mathcal{L} = \zeta_1 \mathcal{L}_{\text{pred}} + \zeta_2 \mathcal{L}_{\text{imp}}. \tag{15}$$

Inference Strategy. During inference, the goal is to predict the correct relation $\hat{r}$ for the given sentence and entity pair. For each candidate relation $r \in R$, the model generates logits for both the positive (l^+) and negative (l^-) propositions. We select the relation that maximizes the contrastive margin between these opposing assertions:

$$\hat{r} = \arg\max_{r\in R} (l^+ - l^-). \tag{16}$$

This strategy effectively enhances robustness by requiring the model to simultaneously affirm the relation and reject its semantic negation.

4 Experiments

4.1 Datasets

Following the majority of prior ZSRE work, we conduct experiments on two widely used datasets: **Wiki-ZSL** (Chen and Li, 2021), a large-scale dataset generated using a distant supervision approach, and **FewRel** (Han et al., 2018), a high-quality, human-annotated dataset. The detailed statistics are provided in **Appendix B**.

Unseen Relation	Methods	Wiki-ZSL			FewRel			Avg.
		P.	*R.*	F_1	*P.*	*R.*	F_1	
m=5	ZS-BERT (Chen and Li, 2021)	71.54	72.39	71.96	76.96	78.86	77.90	74.93
	RE-Matching (Zhao et al., 2023)	78.19	78.41	78.30	92.82	92.34	92.58	85.44
	AlignRE (Li et al., 2024c)	83.11	80.30	81.64	93.30	92.90	93.09	87.37
	EMMA (Li et al., 2024b)	91.32	90.65	90.98	94.87	94.48	94.67	92.83
	CE-DA (Zhang et al., 2025)	88.01	87.02	87.51	95.26	95.08	95.17	91.34
	GLiREL (Boylan et al., 2025)	89.41	80.67	83.28	96.84	93.41	94.20	88.74
	G-NSR (ours)	92.92	93.16	**93.04**	97.53	97.47	**97.50**	**95.27**
m=10	ZS-BERT (Chen and Li, 2021)	60.51	60.98	60.74	56.92	57.59	57.25	59.00
	RE-Matching (Zhao et al., 2023)	74.39	73.54	73.96	83.21	82.64	82.93	78.45
	AlignRE (Li et al., 2024c)	75.00	73.26	74.10	86.41	85.14	85.75	79.93
	EMMA (Li et al., 2024b)	86.00	84.55	85.27	87.97	86.48	87.22	86.25
	CE-DA (Zhang et al., 2025)	82.54	81.82	82.16	88.61	87.60	88.10	85.13
	GLiREL (Boylan et al., 2025)	89.87	81.56	83.67	91.09	87.42	87.60	85.64
	G-NSR (ours)	89.88	89.88	**89.88**	91.59	90.85	**91.22**	**90.55**
m=15	ZS-BERT (Chen and Li, 2021)	34.12	34.38	34.25	35.54	38.19	36.82	35.54
	RE-Matching (Zhao et al., 2023)	67.31	67.33	67.32	73.80	73.52	73.66	70.49
	ZS-SKA (Gong and Eldardiry, 2024)	41.78	40.50	39.30	45.03	51.86	46.99	43.15
	AlignRE (Li et al., 2024c)	69.01	67.52	68.26	77.63	77.00	77.31	72.79
	EMMA (Li et al., 2024b)	78.51	77.63	78.07	80.47	79.73	80.10	79.09
	CE-DA (Zhang et al., 2025)	75.33	74.94	75.13	84.03	82.60	83.31	79.22
	GLiREL (Boylan et al., 2025)	79.44	74.81	73.91	88.14	84.69	84.48	79.20
	G-NSR (ours)	82.31	81.12	**81.71**	88.58	87.62	**88.10**	**84.91**

Table 2: Comparison results of the proposed G-NSR on Wiki-ZSL and FewRel datasets. **Bold** marks the highest score, underline marks the second-hightest score. **Avg.** denotes the average of all the F1-Scores on two datasets. All baseline results are sourced from the original papers.

Consistent with prior work (Chen and Li, 2021), the number of unseen relations is set to $m \in \{5, 10, 15\}$. The data is partitioned using five random seeds, producing five folds, and the average performance is reported.

4.2 Experimental Settings and Metrics

In experiments, we employ BERT-base (Devlin et al., 2019) as the backbone model. All experiments are conducted on a 24GB NVIDIA RTX 4090 GPU. The maximum number of candidate relations in training phase is set to 4, while the weights ζ_1 for the Prediction Loss $\mathcal{L}_{\text{pred}}$ and ζ_2 for the Implication Loss $\mathcal{L}_{\text{imp}}$ are set to 0.5 and 1.0, respectively. A detailed experimental settings and the analysis of hyperparameters (ζ_1 and ζ_2) are provided in **Appendix C**.

For the evaluation metrics, we use the macro Precision (P.), Recall (R.) and F1-Score (F_1) to be consistent with previous work (Chen and Li, 2021).

4.3 Baselines

We compare our method with a comprehensive set of strong baselines: ZS-BERT (Chen and Li, 2021), RelationPrompt (Chia et al., 2022), RE-Matching (Zhao et al., 2023), SUMASK (Li et al., 2023), ChatIE (Wei et al., 2023), ZS-SKA (Gong and Eldardiry, 2024), MICRE (Li et al., 2024a), AlignRE (Li et al., 2024c), EMMA (Li et al., 2024b), RE-GAR-AD (Li et al., 2025), CE-DA (Zhang et al., 2025), and GLiREL (Boylan et al., 2025).

5 Results and Analysis

5.1 Main Results

The results in Table 2 highlight the superiority of G-NSR across both datasets. This advantage is particularly pronounced in the challenging $m = 15$ setting, where our model achieves F1-scores of **81.71** on Wiki-ZSL and **88.10** on FewRel. The average F1-score of **84.91** represents a lead of **5.69** points over the strongest baseline. Moreover, G-NSR establishes new state-of-the-art results across all settings ($m \in \{5, 10, 15\}$), demonstrating its stable generalization capability.

5.2 Comparison with LLMs-based Methods

Given the prominent capabilities of large language models (LLMs) in zero-shot learning, we also conducted a comparative analysis of our model against a number of representative LLM-based baselines. Table 3 summarizes the performance of these baselines, alongside the parameter sizes of their respective backbone models. The results clearly

Methods	Backbone Params	Wiki-ZSL			FewRel			Avg.
		m=5	m=10	m=15	m=5	m=10	m=15	
RelationPrompt$_{BART\&GPT-2}$ (Chia et al., 2022)	264M	76.63	71.50	65.74	89.30	79.96	73.40	76.09
ChatIE$_{Qwen3}$ (Wei et al., 2023)†	235B	77.24	71.48	66.11	90.37	82.24	75.09	77.09
ChatIE$_{GPT-5}$ (Wei et al., 2023)†	-	82.12	74.38	67.19	93.14	85.07	76.09	79.67
MICRE$_{T5-3B}$ (Li et al., 2024a)	3B	76.03	71.30	64.76	88.99	79.67	70.98	75.29
MICRE$_{LLaMA}$ (Li et al., 2024a)	7B	77.48	73.60	67.99	90.59	81.48	74.77	77.65
RE-GAR-AD$_{GPT-3.5-turbo}$ (Li et al., 2025)	-	88.63	77.70	70.02	92.59	82.78	76.47	81.37
G-NSR$_{BERT-base}$ (ours)	110M	**93.04**	**89.88**	**81.71**	**97.50**	**91.22**	**88.10**	**90.24**

Table 3: Comparison results with *LLMs-based methods*. We report the backbone models used in their baselines along with the number of parameters (the parameter size of GPT series is not publicly disclosed). **Avg.** denotes the average of F1-Scores in six settings. † are our reproduction, using the settings provided in the original paper.

Methods	FewRel		
	$P.$	$R.$	F_1
G-NSR	**88.58**	**87.62**	**88.10**
- **w/o.** Duality	84.46	82.92	83.67
- **w/o.** Type Constraints	85.06	82.90	83.96
- **w/o.** Implication Loss	86.33	84.13	85.21
- **w/o.** Neuro-Symbolic Reasoner	85.06	83.52	84.28
- **w/o.** Local Perspective	84.01	82.26	83.12

Table 4: Ablation studies of G-NSR under m=15.

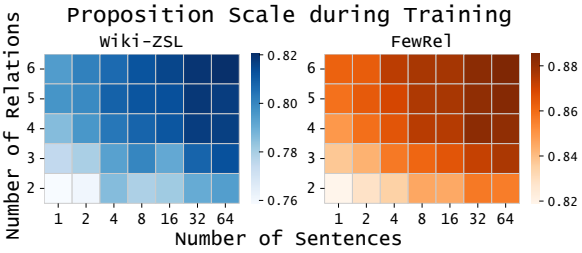

Figure 3: Impact of global proposition scale during training on Wiki-ZSL and FewRel datasets.

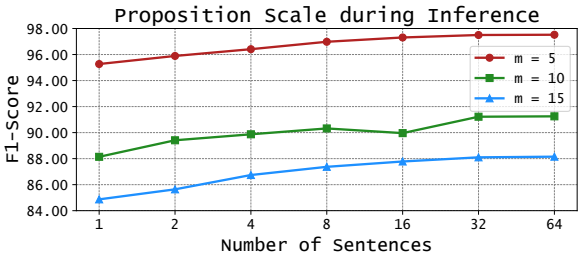

Figure 4: Impact of global proposition scale during inference on FewRel dataset.

demonstrate that our proposed method consistently achieves superior performance over these LLM-based approaches while being substantially more parameter-efficient.

5.3 Ablation Studies

Table 4 summarizes the ablation results. *First*, regarding schema design for constructing propositions, removing negative propositions (Duality) or excluding entity type constraints (Type Constraints) impairs performance, confirming that high-quality, contrastive inputs are foundational. *Second*, regarding the reasoning mechanism, training without the global module (Neuro-Symbolic Reasoner) or the auxiliary logical supervision (Implication Loss) leads to significant declines, validating the necessity of logical consistency reasoning across global predictions. Most critically, the variant without Local Perspective—which discards the intrinsic sentence matching score to rely solely on aggregated global evidence—yields the lowest performance. This indicates that well-grounded local semantics enable global reasoning to be more effective.

5.4 Impact of Global Proposition Scale

We verify that the model performance improves as the scale of global proposition increases—specifically, the number of sentences (N) and candidate relations (M) in Figure 2. We analyze this trend in both training and inference.

Scale during Training. Figure 3 shows a clear positive trend: performance consistently gets better as both numbers increase. Specifically, a larger N enables effective computation of the Implication Loss by providing necessary logical antecedents, while more candidate relations M help the model better distinguish the correct relation from others.

Scale during Inference. Figure 4 shows a consistent upward trend. By processing more sentences together, the model uses high-confidence predictions to help clarify ambiguous ones. Notably, **even with a single sentence** ($N = 1$), i.e., **aligning with the local-perspective reasoning paradigm of existing methods**, our method still outperforms baseline models. This is because there

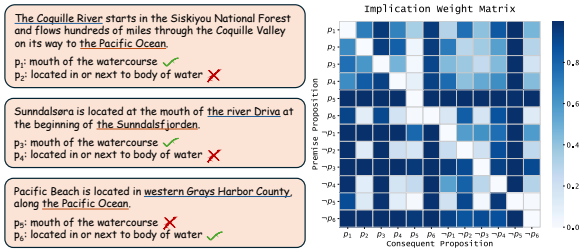

Figure 5: Visualization of implication weight matrix for three representative sentences, illustrating the global reasoning mechanism.

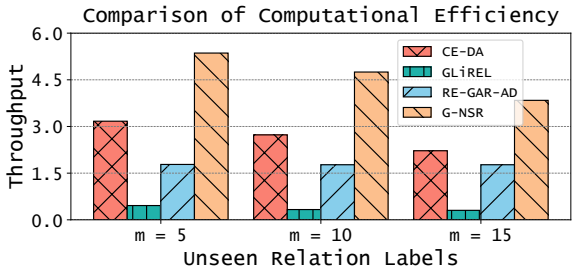

Figure 6: Comparison of inference throughput on FewRel across varying numbers of unseen relation labels (higher throughput indicates faster inference).

are still multiple candidate relations for that single sentence, which allows the model to perform reasoning by analyzing the logical dependencies within the candidate set itself.

However, we observe that the performance gains become less significant as the numbers continue to grow, we set $M = 4$ for training phase and $N = 32$ for both phases to balance accuracy and speed.

5.5 Analysis on Duality Type-Constrained Relation Schema Configurations

To investigate how the format of propositions affects reasoning performance, we compare our Duality Type-Constrained Relation Schemas against three simplified configurations. The results are presented in **Appendix D**. We observe that the baseline configuration using raw relation labels already yields reasonable performance. Moreover, incorporating Duality and Type Constraints leads to further consistent improvements, with the full configuration achieving the best results across all settings. This confirms that higher-quality inputs allow the Neuro-Symbolic Reasoner, as the core engine, to play a larger role. Specifically, type constraints filter entity mismatches, while duality sharpens decision boundaries. This structured foundation reduces semantic ambiguity, facilitating more accurate global reasoning.

5.6 Analysis on Global Reasoning Mechanism

To illustrate the global reasoning mechanism, we visualize the implication weights for three representative sentences in Figure 5.

Initially, the model exhibits uncertainty regarding the first two ambiguous sentences, indicated by moderate weights in their corresponding submatrices. In contrast, the third sentence acts as a high-confidence anchor. The model correctly identifies its false propositions (e.g., $p_5, \neg p_6$), assigning them uniformly high implication weights to other propositions—a behavior consistent with the logical Principle of Explosion (a false premise implies any conclusion). Crucially, the anchor propagates corrective evidence to the ambiguous instances: the truth of p_6 strongly implies the negation of incorrect options for the first two sentences (e.g., $p_6 \rightarrow \neg p_2$). This demonstrates how the reasoner leverages confident judgments to resolve local ambiguities, ensuring global logical consistency.

Several additional **case studies** comparing prediction probabilities under the local perspective and global reasoning are provided in **Appendix E**.

5.7 Analysis on Computational Efficiency

We compare the inference throughput (i.e., the number of sentences processed per second) of our framework against representative baselines on the FewRel dataset. As shown in Figure 6, throughput naturally decreases as the proposition scale increases. Nonetheless, our method G-NSR consistently maintains the highest processing speed. Even in the demanding $m = 15$ setting, it retains a significant advantage over baselines, demonstrating that our approach achieves high performance with low computational overhead.

6 Conclusion

We present G-NSR, a neuro-symbolic framework for ZSRE that innovatively formulates the ZSRE task as a set of propositions and performs global reasoning to produce logically consistent predictions. G-NSR constructs a global set of positive and negative propositions from input sentences and relations via Duality Type-Constrained Relation Schemas, and then uses a core Neuro-Symbolic Reasoner to model logical dependencies (e.g., implication and disjunction) and refine proposition truth values based on collective evidence. Extensive experiments show that G-NSR consistently outperforms prior methods and achieves new state-

of-the-art results. We hope this work broadens the applicability of neuro-symbolic reasoning to zero-shot relation extraction and offers useful insights for future relation extraction research.

Limitations

Our framework is designed to prioritize rigorous logical consistency through global reasoning. Theoretically, the throughput naturally decreases as the scale of propositions grows. However, since our reasoning mechanism is fundamentally formulated as efficient matrix operations, it is inherently parallel-friendly. As demonstrated in Figure 6, G-NSR maintains the highest processing speed among all baselines even under complex settings. Thus, this computational characteristic does not constitute a practical bottleneck; rather, it represents a strategic trade-off that secures significant accuracy gains with negligible overhead. Future work may explore advanced optimization strategies to further extend this efficiency to massive-scale scenarios.

In addition, as discussed in Sections 5.4 and 5.5, the effectiveness of our neuro-symbolic ZSRE framework may be affected, to some extent, by the number of constructed propositions and the relation schema configurations used to construct them. While we observe consistent robustness across varying proposition scales and schema configurations on two widely used ZSRE benchmarks, it remains valuable to further assess these factors in broader and more diverse real-world application scenarios.

Acknowledgements

The authors sincerely thank the anonymous reviewers for their valuable comments and suggestions, which have greatly improved this paper. This work is supported by the National Natural Science Foundation of China (62276057).

References

Samy Badreddine, Artur d'Avila Garcez, Luciano Serafini, and Michael Spranger. 2022. Logic tensor networks. *Artificial Intelligence*, 303:103649.

Antoine Bosselut, Ronan Le Bras, and Yejin Choi. 2021. Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. In *Proceedings of the AAAI conference on Artificial Intelligence (AAAI)*, pages 4923–4931.

Jack Boylan, Chris Hokamp, and Demian Gholipour Ghalandari. 2025. GLiREL - generalist model for zero-shot relation extraction. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pages 8230–8245.

Chih-Yao Chen and Cheng-Te Li. 2021. ZS-BERT: Towards zero-shot relation extraction with attribute representation learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pages 3470–3479.

Yew Ken Chia, Lidong Bing, Soujanya Poria, and Luo Si. 2022. RelationPrompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction. In *Findings of the Association for Computational Linguistics (ACL)*, pages 45–57.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pages 4171–4186.

Richard Evans and Edward Grefenstette. 2018. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61:1–64.

Jiaying Gong and Hoda Eldardiry. 2024. Prompt-based zero-shot relation extraction with semantic knowledge augmentation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (COLING)*, pages 13143–13156.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4803–4809.

Guozheng Li, Peng Wang, and Wenjun Ke. 2023. Revisiting large language models as zero-shot relation extractors. In *Findings of the Association for Computational Linguistics (EMNLP)*, pages 6877–6892.

Guozheng Li, Peng Wang, Jiajun Liu, Yikai Guo, Ke Ji, Ziyu Shang, and Zijie Xu. 2024a. Meta in-context learning makes large language models better zero and few-shot relation extractors. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 6350–6358.

Shilong Li, Ge Bai, Zhang Zhang, Ying Liu, Chenji Lu, Daichi Guo, Ruifang Liu, and Sun Yong. 2024b. Fusion makes perfection: An efficient multi-grained matching approach for zero-shot relation extraction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pages 79–85.

Tao Li and Vivek Srikumar. 2019. Augmenting neural networks with first-order logic. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 292–302.

Zehan Li, Fu Zhang, and Jingwei Cheng. 2024c. Alignre: An encoding and semantic alignment approach for zero-shot relation extraction. In *Findings of the Association for Computational Linguistics (ACL)*, pages 2957–2966.

Zehan Li, Fu Zhang, Tianyue Peng, He Liu, and Jingwei Cheng. 2025. Generation-augmented retrieval: Rethinking the role of large language models in zero-shot relation extraction. In *Findings of the Association for Computational Linguistics (EMNLP)*, pages 17928–17941.

Liangming Pan, Alon Albalak, Xinyi Wang, and William Wang. 2023. Logic-lm: Empowering large language models with symbolic solvers for faithful logical reasoning. In *Findings of the Association for Computational Linguistics (EMNLP)*, pages 3806–3824.

Tim Rocktäschel and Sebastian Riedel. 2017. End-to-end differentiable proving. *Advances in neural information processing systems (NeurIPS)*, 30.

Somin Wadhwa, Silvio Amir, and Byron C Wallace. 2023. Revisiting relation extraction in the era of large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 15566–15589.

Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, and 1 others. 2023. Chatie: Zero-shot information extraction via chatting with chatgpt. *arXiv preprint arXiv:2302.10205*.

Michihiro Yasunaga, Xinyun Chen, Yujia Li, Panupong Pasupat, Jure Leskovec, Percy Liang, Ed H Chi, and Denny Zhou. 2024. Large language models as analogical reasoners. In *The Twelfth International Conference on Learning Representations (ICLR)*.

Urchade Zaratiana, Nadi Tomeh, Pierre Holat, and Thierry Charnois. 2024. An autoregressive text-to-graph framework for joint entity and relation extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 19477–19487.

Fu Zhang, He Liu, Zehan Li, and Jingwei Cheng. 2025. CE-DA: Custom embedding and dynamic aggregation for zero-shot relation extraction. In *Proceedings of the 31st International Conference on Computational Linguistics (COLING)*, pages 9814–9823.

Jun Zhao, WenYu Zhan, Xin Zhao, Qi Zhang, Tao Gui, Zhongyu Wei, Junzhe Wang, Minlong Peng, and Mingming Sun. 2023. RE-matching: A fine-grained semantic matching method for zero-shot relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 6680–6691.

A Detailed Network Structure Used in Proposition Encoder

A.1 Details of the EntityRep module

The `EntityRep` module, mentioned in Proposition Encoder of Section 3.4, is designed to generate a comprehensive entity representation by capturing the rich interactions between the entity's start and end boundary embeddings. Let $\mathbf{h}_{start}, \mathbf{h}_{end} \in \mathbb{R}^d$ denote the input embeddings corresponding to the start and end tokens of an entity (e.g., $\mathbf{h}_{[\mathrm{H}]}$ and $\mathbf{h}_{[\mathrm{/H}]}$ for the head entity). The module operates in three steps:

Multi-View Feature Fusion: To capture different aspects of the entity span, we compute four distinct feature vectors: the original start and end embeddings, their element-wise product (capturing similarity), and their difference (capturing directionality). These features are concatenated to form a composite vector $\mathbf{z} \in \mathbb{R}^{4d}$:

$$\mathbf{z} = [\mathbf{h}_{start}; \mathbf{h}_{end}; \mathbf{h}_{start} \odot \mathbf{h}_{end}; \mathbf{h}_{start} - \mathbf{h}_{end}], \tag{17}$$

where $\odot$ denotes element-wise multiplication and $[\cdot;\cdot]$ denotes concatenation.

Non-Linear Projection: The composite vector $\mathbf{z}$ is projected back to the original dimension d via a linear layer, followed by a GELU activation and Dropout regularization:

$$\mathbf{h}_{proj} = \text{Dropout}(\text{GELU}(\mathbf{W}_e\mathbf{z} + \mathbf{b}_e)), \tag{18}$$

where $\mathbf{W}_e \in \mathbb{R}^{d\times 4d}$ and $\mathbf{b}_e \in \mathbb{R}^d$ are learnable parameters.

Residual Connection and Normalization: To facilitate gradient flow and stabilize training, we employ a residual connection. The residual term is defined as the average of the start and end embeddings. The final output $\mathbf{e}$ is obtained by adding this residual to the projected features and applying Layer Normalization:

$$\mathbf{e} = \text{LayerNorm}\left(\mathbf{h}_{proj} + \frac{\mathbf{h}_{start} + \mathbf{h}_{end}}{2}\right). \tag{19}$$

This architecture ensures that the final representation $\mathbf{e}$ retains the original boundary information while incorporating complex interaction features.

A.2 Details of the Fusion layer

The `Fusion` layer, also mentioned in Proposition Encoder of Section 3.4, is responsible for integrating the head and tail representations into a unified

vector. We design a Gated Residual Network to perform this integration, which is applied in parallel to both the sentence-specific entities and the schema-specific placeholders.

First, we construct the input vectors by concatenating the head and tail representations for the sentence (s) and the schema (τ) respectively. Let d be the hidden dimension of the entity representations. The inputs $\mathbf{x}_s, \mathbf{x}_\tau \in \mathbb{R}^{2d}$ are defined as:

$$\mathbf{x}_s = [\mathbf{e}_h^{(s)}; \mathbf{e}_t^{(s)}], \quad \mathbf{x}_\tau = [\mathbf{e}_h^{(\tau)}; \mathbf{e}_t^{(\tau)}]. \tag{20}$$

Then, we define a non-linear transformation function $\Phi(\mathbf{x})$, corresponding to the `Fusion` class in our implementation. This function processes an input vector $\mathbf{x}$ through a gating mechanism followed by a residual refinement block. It consists of three Feed-Forward Networks (FFN), denoted as F_1, F_2, F_3. The computation proceeds as follows:

Contextual Gating: We generate a gating vector $\mathbf{g}$ and a candidate feature vector $\tilde{\mathbf{x}}$. The gate controls the information flow using a Sigmoid activation:

$$\mathbf{g} = \sigma(F_1(\mathbf{x})), \quad \tilde{\mathbf{x}} = F_2(\mathbf{x}). \tag{21}$$

The intermediate representation $\mathbf{m}$ is obtained via element-wise multiplication:

$$\mathbf{m} = \mathbf{g} \odot \tilde{\mathbf{x}}. \tag{22}$$

Residual Refinement: To further refine the features while preserving gradient stability, we process $\mathbf{m}$ through a third FFN and apply a residual connection with Layer Normalization:

$$\mathbf{y} = \text{LayerNorm}(\mathbf{m} + \text{Dropout}(F_3(\mathbf{m}))). \tag{23}$$

Finally, we apply this transformation Φ independently to both the sentence and schema inputs to obtain their fused representations. The final proposition embedding $\mathbf{v}_{p_i} \in \mathbb{R}^{4d}$ is the concatenation of these two outputs:

$$\mathbf{v}_{p_i} = [\Phi(\mathbf{x}_s); \Phi(\mathbf{x}_\tau)]. \tag{24}$$

B Dataset Statistics

The detailed statistics of the two datasets are shown in Table 5.

Wiki-ZSL (Chen and Li, 2021) is a large-scale dataset specifically designed for the Zero-Shot Learning (ZSL) setting in relation extraction. It is generated from the Wikidata knowledge base using a distant supervision approach. The dataset comprises 113 relations and a total of 94,383 instances.

Dataset	#Instances	#Entities	#Relations	Avg.L
Wiki-ZSL	94,383	77,623	113	24.85
FewRel	56,000	72,954	80	24.95

Table 5: Statistics of two datasets. *Avg.L*: the average number of words in each instance (i.e., each sentence).

FewRel (Han et al., 2018) is a high-quality, human-annotated dataset that has become a standard benchmark for ZSRE. It is collected from Wikipedia and manually verified by crowd workers. The dataset contains 80 relations in total, each with 700 instances, resulting in 56,000 instances overall.

Hyperparameter	**Value**
BERT	base
Optimizer	AdamW
Encoder Learning rate	5×10^{-5}
Others Learning rate	1×10^{-4}
Batch Size	32
Epoch	3
Warmup Ratio	10%
Dropout	0.1
Maximum Relations	4
ζ_1	0.5
ζ_2	1.0

Table 6: Hyperparameter setting.

C Detailed Experimental Settings and Hyperparameter Studies

C.1 Detailed Experimental Settings

In Table 6, we provide the detailed hyperparameter values used in our method. We employ BERT-base (Devlin et al., 2019) as the backbone model to evaluate the performance of the proposed method. Model parameters are optimized with the AdamW optimizer, using learning rates of 5×10^{-5} for the encoder and 1×10^{-4} for the remaining components. The batch size is set to 32, and training is conducted for 3 epochs. The initial 10% of training incorporates a warmup phase, where the learning rate increases from 0 to the specified base rate. The maximum number of candidate relations in training phase is set to 4, while the weights ζ_1 for the Prediction Loss $\mathcal{L}_{\text{pred}}$ and ζ_2 for the Implication Loss $\mathcal{L}_{\text{imp}}$ are set to 0.5 and 1.0, respectively. All experiments are implemented using the PyTorch framework and executed on a 24GB NVIDIA RTX 4090 GPU.

Configurations	Examples of Schema Formats	FewRel			Avg.
		m=5	m=10	m=15	
- **w/o.** Duality	[TAIL] (a publisher ...) was/is the publisher of [HEAD] (a book ...). [TAIL] (not a ...) was not/is not the publisher of [HEAD] (not a ...).	95.01	87.49	83.67	88.72
- **w/o.** Type Constraints	[TAIL] (a publisher ...) was/is the publisher of [HEAD] (a book ...). [TAIL] (not a ...) was not/is not the publisher of [HEAD] (not a ...).	96.57	87.97	83.96	89.50
- **w/o.** Both	[TAIL] (a publisher ...) was/is the publisher of [HEAD] (a book ...). [TAIL] (not a ...) was not/is not the publisher of [HEAD] (not a ...).	94.74	86.29	82.54	87.86
G-NSR	[TAIL] (a publisher ...) was/is the publisher of [HEAD] (a book ...). [TAIL] (not a ...) was not/is not the publisher of [HEAD] (not a ...).	**97.50**	**91.22**	**88.10**	**92.27**

Table 7: Performance comparison of different schema configurations on the FewRel dataset. The gray parts represent the removed content. "w/o. Both" denotes the baseline configuration using raw relation labels.

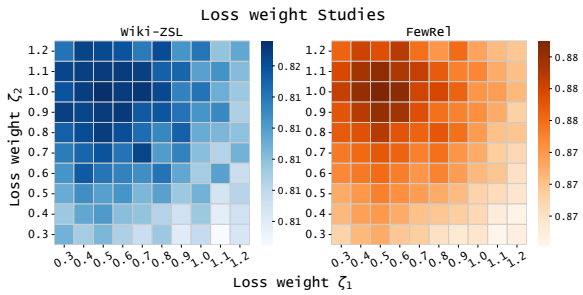

Figure 7: Impact of the loss weights ζ_1 and ζ_2 on performance.

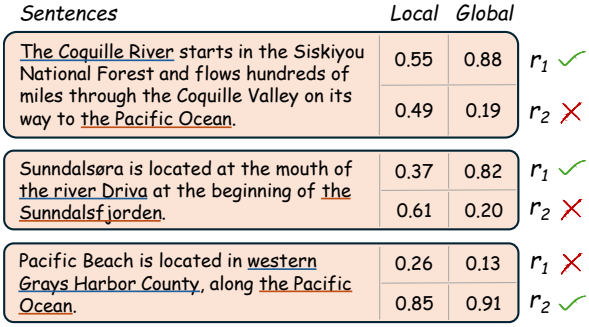

Figure 8: Case study comparing prediction probabilities under Local Perspective and Global Reasoning. r_1 denotes "*mouth of the watercourse*" and r_2 denotes "*located in or next to body of water*".

C.2 Hyperparameter Studies

As observed in Figure 7, the model achieves optimal performance when ζ_1 is set to 0.5 and ζ_2 to 1.0, with a gradual decline as the values deviate from this center. This phenomenon arises because the Implication Loss (ζ_2) serves as a critical logical regularizer. Assigning a relatively higher weight to ζ_2 ensures that the intermediate reasoning adheres strictly to logical rules, providing a robust foundation for the final prediction. Conversely, the smooth decay in performance towards the surrounding regions indicates that the framework maintains stability and is not overly sensitive to minor hyperparameter fluctuations. Consequently, we apply the optimal weights of $\zeta_1 = 0.5$ and $\zeta_2 = 1.0$ across all experiments.

D Performance Comparison of Different Relation Schema Configurations

We investigate how the format of propositions affects reasoning performance, we compare our Duality Type-Constrained Relation Schemas against three simplified configurations, as mentioned in Section 5.5. The results are presented in Table 7.

The results show that the baseline configuration using raw relation labels (w/o Both) already yields reasonable performance. Moreover, incorporating Duality and Type Constraints leads to further consistent improvements, with the full configuration achieving the best results across all settings. These findings demonstrate the effectiveness and generalizability of constructing propositions with different relation schema configurations, and highlight that our proposed Duality Type-Constrained formulation delivers the strongest performance.

E Case Study

Figure 8 compares the predicted probabilities for two similar relations: "*mouth of the watercourse*" (r_1) and "*located in or next to body of water*" (r_2). Operating from a **local perspective**, the independent processing leads to confusion. For the first sentence, the model produces ambiguous scores. More critically, in the second sentence, the phrase "*is located at*" overlaps with the definition of r_2. This misleads the model into incorrectly predicting r_2 (score 0.61) instead of the correct relation r_1.

In contrast, **global reasoning** effectively resolves these issues by utilizing cross-sentence context. The third sentence serves as a clear, high-confidence example of r_2. Leveraging this evidence, the model better distinguishes the boundaries between the two relations. Consequently, it corrects the prediction for the second sentence, raising the score of the correct relation r_1 to 0.82. Additionally, the confidence for the first sentence is significantly boosted from 0.55 to 0.88, demonstrating the robustness of our framework.