Spatial Relation Classification on Few-shot In-context Learning

Guoqi Yang, Peifeng Li, Qiaoming Zhu
School of Computer Science and Technology, Soochow University, Suzhou, China
20224227046@stu.suda.edu.cn, {pfli,qmzhu}@suda.edu.cn

***Abstract*—The goal of the spatial relation classification task is to identify the type of spatial relations between entities extracted from text. Although large language models (LLMs) have demonstrated the potential to achieve excellent performance across various tasks through In-Context Learning (ICL), there remains a lack of exploration in the domain of spatial relation classification. The gap between ICL and fully supervised training models primarily stems from two aspects: current example retrieval methods fall short in capturing the semantic relevance of spatial elements and spatial relations; and the absence of explicit explanations for the mapping between inputs and spatial relation labels in examples hinders the model's ability to effectively leverage contextual information.To address these challenges, we propose the Spatial-ICL method, which specifically optimizes the application of ICL in spatial relation classification. Experimental results demonstrate that Spatial-ICL outperforms the ICL baselines on GPT-4 in terms of performance.**

***Index Terms*—Spatial relation classification, Large Language Models, In-Context Learning**

I. Introduction

Spatial relation classification concentrates on identifying spatial semantic links expressed in text, laying the groundwork for scene comprehension and spatial reasoning. For example, by processing textual data, a system can recognize relations between individuals and places or objects and environments, such as "adjacent to," "inside," or "below." Classifying these relations is crucial for applications like knowledge graph development [1], and question-answering systems [2].

Spatial relations generally involve spatial elements and spatial relation triggers. Spatial elements are entities or events with spatial characteristics, while spatial relation triggers are linking words that signal spatial relations. Typically, a spatial relation in text comprises three parts: the spatial relation subject, the spatial trigger, and the spatial relation object. Both the subject and object are subtypes of spatial elements. The subject can be either a trajector or a mover, and the object can be a landmark or a goal, depending on the type of spatial relations. Figure 1 provides an example of spatial relations within a sentence. In Figure 1, the text contains three spatial relations: QSLINK represents a topological spatial relation, OLINK denotes a directional spatial relation, and MOVELINK indicates a dynamic spatial relation.

Prior studies [3] have attempted to apply ICL to relation classification tasks in specific domains, but results indicate a significant performance gap compared to fine-tuned models. This suggests that current ICL strategies have yet to fully unlock the potential of LLMs in spatial relation classification, with performance constrained by the effectiveness of example selection and the model's depth of contextual understanding.

Sentence:
The ancient [ruins] are [hidden] by [vines], and you might be wondering why [she] chose to [trek] to the [center] [of] [Mexico].

Relation:

{ QSLINK:	{ OLINK:	{ MOVELINK:
trajector = center	trajector = vines	mover = she
trigger = of	trigger = hidden	trigger = trek
landmark = Mexico }	landmark = ruins }	goal = center }

Fig. 1: An example of a spatial relation triplet.

When LLMs employ ICL in spatial relation classification tasks, their suboptimal performance can be attributed to two primary factors. First, example retrieval lacks sufficient relevance to spatial elements and spatial relations [4]. ICL typically relies on random selection or k-nearest neighbor (kNN) retrieval methods based on sentence embeddings [3], [5] to obtain examples. However, such sentence embedding-based retrieval focuses primarily on overall semantic similarity, failing to adequately capture the features of spatial elements and spatial relations within sentences. Second, examples lack interpretability regarding the mapping process between inputs and spatial relations [6]. Traditional ICL methods typically present examples in the form of input-label pairs, without further elaboration on the reasoning process behind spatial relations. This can lead LLMs to over-rely on superficial linguistic features for predictions, overlooking the diversity and semantic complexity of spatial relation expressions.

To address these issues, we propose Spatial-ICL, a method designed to optimize ICL for spatial relation classification. Our approach introduces a task-oriented retrieval mechanism that incorporates spatial element information and uses fine-tuned representations to improve example relevance. Additionally, we enhance examples with gold-label-guided reasoning to support deeper semantic understanding. Experimental results on the SpaceEval dataset show that Spatial-ICL outperforms existing ICL baselines on GPT-4, demonstrating its effectiveness in few-shot settings.

II. Related Work

Task 8 (SpaceEval) of SemEval 2015 [7], building on the SpRL tasks from SemEval 2012 and SemEval 2013, was

further expanded into a comprehensive information extraction and classification task. It adopts the ISO-Space standard [8], enhancing the semantics of SpRL through finer granularity and extending the data domains. SpaceEval aims to identify and classify spatial information in text, encompassing spatial elements such as place, spatial entity, path, and motion, as well as spatial relations (e.g., QSLINK, OLINK, etc.).

Research methods for spatial relation classification can primarily be divided into two categories: traditional machine learning approaches and deep learning methods based on neural networks. With the rapid advancement of deep learning, neural network-based methods have gradually become the mainstream in spatial relation classification research.

Nichols et al. [9] pioneered the use of a CRF model integrated with distributed word representations and lexical-syntactic features to identify spatial elements and signals, followed by a separate multi-class classification model that employs syntactic and semantic features to concurrently determine relation types and assign spatial roles. D'Souza et al. [10] developed a MOVELINK spatial relation extraction approach using sieves and extended syntax trees, combining multi-channel sieves and tree kernel methods to progressively detect spatial elements and their roles in complex MOVELINK relations, leveraging earlier sieve outputs to inform subsequent decisions. Salaberri et al. [11] utilized a four-stage pipeline approach based on WordNet and SVM, adhering to the ISO-Space framework to automatically extract spatial elements, signals, and relations from text.

Ramrakhiyani et al. [12] presented a simple two-step neural network method for spatial role labeling, first training a BiLSTM for sequence labeling to produce contextual word embeddings, then using these embeddings alongside dependency parsing to generate candidate relations, followed by another neural network to classify static spatial relations and their roles. Shin et al. [13] introduced a BERT-based model for spatial information extraction, splitting the task into two subtasks: spatial element extraction and spatial relation classification. They applied BERT+CRF for joint extraction of spatial elements and roles, generated triplets using a triplet candidate generator, and used an enhanced R-BERT [14] for triplet relation classification. Wang et al. [15] proposed the HMCGR hybrid model, which combines the strengths of classification and generation models and incorporates a relation symmetry-based evaluation mechanism to improve spatial relation extraction performance. Yang et al. [16] introduced a spatial relation classification model enhanced by AMR graphs, boosting semantic discrimination by adding special markers and incorporating an AMR graph aggregation module to capture semantic interactions and structural information between sentences.

III. Methodology

In this section, we first elaborate on the prompt design method in Spatial-ICL, formalizing the spatial relation classification task as a language generation problem. In the candidate triplet generation phase, we initially employ the BERT-CRF model to identify spatial roles in the text, followed by permuting spatial elements to obtain a set of candidate triplets. In the optimized retrieval phase, we leverage the identified spatial role labels to refine the retrieval process, thereby enhancing the sample selection quality for ICL. Next, we introduce two core optimization modules to improve the effectiveness of ICL in spatial relation classification: task-oriented example retrieval and reasoning guided by gold labels. Figure 2 illustrates the specific workflow for processing test inputs.

A. Prompt Construction

This method designs a prompt for each test sample and feeds it into the GPT-4 model for processing. Each prompt comprises the following core components:

1) Task Description: We begin with a concise overview of the spatial relation classification task and the predefined set of spatial relation types, $\Omega = \{$*QSLINK*, *OLINK*, *MOVELINK*$\}$. The model is explicitly required to output a spatial relation from this predefined set or return NULL if no clear relation exists, where $\mathcal{I}$ denotes the task description instruction.

2) ICL Example Set: We first obtain a K-shot example set through a task-oriented example selector. Then, for each example (c_i, ω_i), we introduce reasoning guidance r_i based on the gold label, forming an enhanced example set (c_i, ω_i, r_i), denoted as $\mathcal{D}$.

3) Test Input: Similar to the examples, we provide the test input x_{test} and expect the LLM to generate the corresponding spatial relation ω_{test} based on the context. In summary, the spatial relation model based on few-shot in-context learning can be expressed as:

$$p\left(\omega_{test} \in \Omega \cup \{NULL\} | \mathcal{I}, \mathcal{D}, x_{test}\right) \tag{1}$$

This expression describes the probability that the model predicts ω_{test} as belonging to the predefined set Ω or NULL, given the task description $\mathcal{I}$, example set $\mathcal{D}$, and test input x_{test}.

B. Task-Oriented Example Retrieval

The examples more similar to the test sample in ICL can lead to more consistent and robust performance [5]. Recent studies often adopt the kNN method to retrieve examples from the training set that are semantically similar to the test input, serving as contextual support for few-shot learning. However, the effectiveness of kNN retrieval heavily depends on the quality of the embedding space representation, typically achieved by encoding test inputs and training examples using pre-trained language models or improved sentence embeddings. The core of spatial relation classification lies in identifying specific spatial connections between triplets of spatial elements, whereas sentence embeddings primarily reflect overall semantic similarity, diverging from the task's focus.

To address this issue, we propose a task-oriented example selection optimization method to enhance the relevance and quality of example retrieval, ensuring that the selected examples are highly pertinent to the spatial relation classification task. This method incorporates the following two optimization strategies:

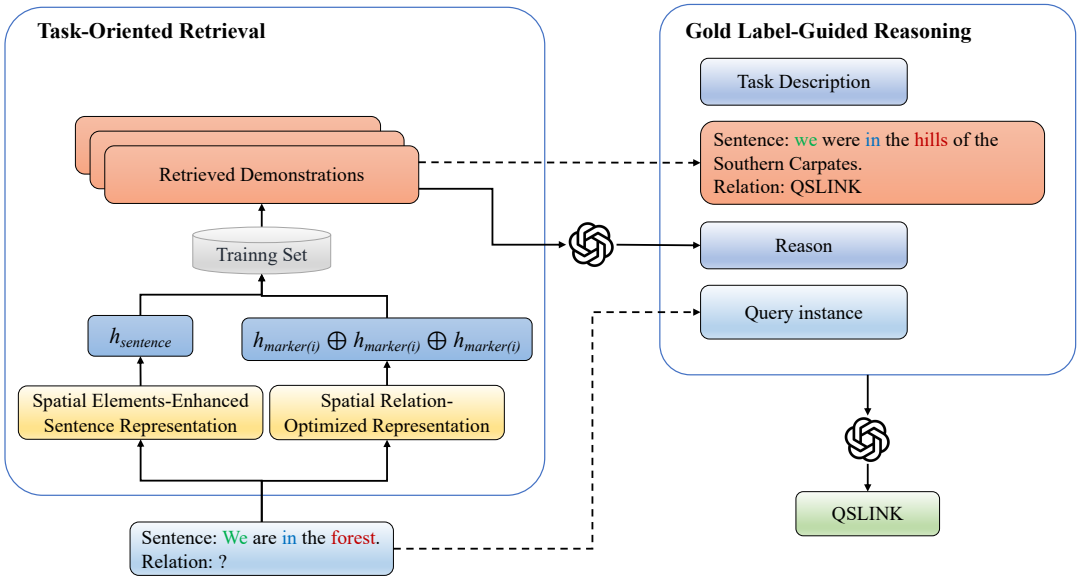

Fig. 2: Overview of our model.

1) Spatial Elements-Enhanced Sentence Representation: Traditional sentence embedding methods typically focus on overall semantics while overlooking key elements specific to a given task. Considering the central role of spatial element triplets in spatial relation identification, this section proposes improving embedding effectiveness by explicitly incorporating spatial element triplet information into sentences. This approach enhances the model's focus on spatial elements and their contexts, thereby improving retrieval performance.

For instance, given the original sentence "He drive again into the hardwood forest," the reconstructed one after spatial element prompting becomes: "The relation between 'He', 'drive', and 'forest' in the context: He drive again into the hardwood forest." This method not only preserves the overall semantics of the sentence during retrieval but also emphasizes spatial information centered on spatial elements, thus increasing the relevance and precision of examples to the task objectives. We adopt the high-performing SimCSE model [17] to compute similarity based on sentence embeddings, ensuring the effectiveness and stability of the representation approach.

2) Spatial Relation-Optimized Representation: Compared to the strategy of incorporating spatial element triplet information into contextual sentences, a more direct and effective optimization approach involves extracting relation representations from a model fine-tuned for spatial relation tasks. By leveraging a representation model fine-tuned specifically for spatial relations, the ability to capture spatial relation features can be further strengthened, thereby enhancing the specificity and quality of example retrieval.

Current BERT-based relation classification methods highlight entities and their types by inserting markers into sentences [18], [19]. In spatial relation classification, adding typed textual markers before and after triplets [16] reinforces spatial role information and enhances semantic differentiation, effectively integrating contextual information with spatial element characteristics. Specifically, for the example sentence "He drive again into the hardwood forest," the input format for the model is adjusted to: "[CLS] $\langle S : trajector\rangle$ He $\langle E : trajector\rangle$ $\langle S : trigger\rangle$ drive $\langle E : trigger\rangle$ again into the hardwood $\langle S : landmark\rangle$ forest $\langle E : landmark\rangle$. [SEP]" Here, $\langle S : trajector\rangle$, $\langle S : trigger\rangle$, and $\langle S : landmark\rangle$ denote markers for spatial elements, explicitly labeling the distinct roles within the spatial relation.

Let the hidden representation of the nth layer of the BERT encoder be h_n, where i, j, and k correspond to the positional indices of the spatial role markers $\langle S : trajector\rangle$, $\langle S : trigger\rangle$, and $\langle S : landmark\rangle$, respectively. The method we propose defines the spatial relation representation as:

$$\text{Spatial-Rel} = h_i \oplus h_j \oplus h_k \tag{2}$$

where $\oplus$ denotes vector concatenation along the primary dimension. Subsequently, this spatial relation representation is fed into a feedforward neural network to predict the probability distribution of spatial relations:

$$p(\omega_{ICL} \in \Omega \cup \{NULL\}|\text{Spatial-Rel}) \tag{3}$$

Through this approach, the model explicitly encodes the different spatial roles within the triplet of a spatial relation, and the resulting relation representation naturally encapsulates rich spatial element information and spatial semantic features, significantly improving representation precision. Moreover, this method effectively compensates for GPT-4's shortcomings in spatial relation classification tasks. Since the spatial relation representation is customized and fine-tuned for the extraction task, the relevance and quality of example retrieval are substantially enhanced.

C. Gold Label-Guided Reasoning

Recent studies on Chain-of-Thought [20], [21] demonstrate that generating reasoning steps can significantly enhance

model performance in commonsense reasoning and numerical reasoning tasks. However, in spatial relation classification tasks, multiple spatial relations may exist between triplets. For example, "library" and "park" could exhibit a relation of proximity or being north of one another. If the reasoning process lacks a clear connection to the gold label, the generated explanations may deviate from the task objectives, leading to reduced classification performance. Therefore, to improve the accuracy of spatial relation classification, we introduce a reasoning guidance strategy based on gold labels. This approach incorporates a reasoning process after example retrieval to provide richer semantic grounding, thereby enhancing the semantic expressiveness of each example.

This section proposes leveraging GPT-4 to generate reasoning processes corresponding to the gold spatial relation labels for each example, thereby enhancing the semantic expressiveness of the examples. First, a query prompt is constructed based on the example: "What clues indicate that the spatial relation between [space element1], [space element2], and [space element3] is [spatial relation] in the sentence [context]?" This prompt is then fed into GPT-4, which is tasked with generating a reasonable reasoning process based on the sentence context and the annotated spatial relation of the triplet. The output format is: "The reasoning is: ..." The generated reasoning evidence focuses primarily on clues related to the spatial element pairs and their annotated relations, ensuring consistency with the objectives of the spatial relation classification task. Finally, the reasoning evidence produced by GPT-4 is combined with the original example to form an enhanced example set.

IV. Experiments

A. Experimental Settings

We evaluate our model on the SpaceEval dataset [7], which is the most popular dataset for spatial relation extraction. Following previous work [15], the training set includes 55 documents annotated with 1872 spatial relations, and the test set consists of 14 documents annotated with 394 spatial relations. For evaluation, we report Precision (P), Recall (R), and Micro-F1 score.

Since there is currently no research specifically targeting few-shot learning for spatial relation classification tasks, we select ICL-Random and ICL-kNN as two baseline systems in this study. Given that our focus is on spatial relation classification, we adopt the AMR-Enhanced model [16] for candidate triplet extraction, enumerating all possible triplets as candidates. We use a dataset of 100 samples to fine-tune previous baseline models. Additionally, we select "gpt-4o" as the model for the Spatial-ICL method and GPT-4-related baselines, setting the maximum input length to 4,097 tokens and employing the same prompt construction approach via the OpenAI API.

B. Results

The implementation details of the two GPT-4 baseline methods are as follows:

ICL-Random: Unlike traditional methods that randomly sample few-shot examples from training data, we introduce a constraint to ensure a balanced label distribution, making the spatial relation category distribution of the selected examples more uniform. This approach outperforms conventional random sampling in terms of performance, making it a more competitive baseline.

ICL-kNN: Existing studies often utilize various sentence embedding techniques for example retrieval. This method employs the SimCSE model, chosen for its superior performance in sentence semantic similarity tasks, as the embedding generation tool in the retrieval phase to select examples from the training set that are semantically similar to the test input.

Fine-Tuned Spatial Relation Classification Model: In the experiments, we adopt AMR-Enhanced model to generate spatial relation representations for supporting example retrieval. This AMR graph-enhanced spatial relation classification model captures spatial semantic features by adding special textual markers for boundaries and spatial roles, making it well-suited for spatial relation classification tasks.

Our Spatial-ICL is compared with existing methods in Table I. Here, Spatial-ICL-SimCSE refers to example retrieval using sentence representations enhanced with spatial elements, while Spatial-ICL-FT denotes retrieval using fine-tuned representations optimized for spatial relations. The results in the table indicate that both Spatial-ICL-SimCSE and Spatial-ICL-FT outperform the sentence embedding-based ICL-kNN by 5.3 and 6.6 points, respectively. This demonstrates that optimizing example selection by incorporating task-relevant semantic information can significantly enhance the performance of spatial relation classification. In a few-shot setting, Spatial-ICL-FT surpasses the fine-tuned baseline models HMCGR and AMR-Enhanced, improving Micro-F1 scores by 14.6 and 10.4, respectively. This suggests that when the retrieval mechanism is equipped with prior knowledge of the spatial relation task, GPT-4's classification capabilities have the potential to exceed those of traditional fine-tuned models.

In the few-shot setting, Spatial-ICL-FT achieves substantial improvements over AMR-Enhanced in F1 scores for the QSLINK and OLINK relations, with gains of 7.5 and 17.2, respectively. These significant enhancements indicate that LLMs offer a stronger advantage in distinguishing these two relations compared to previous methods. The performance improvements for these relations also suggest that our model is not influenced by the number of spatial relations in the training samples. Across smaller subcategories of relations, both Spatial-ICL-SimCSE and Spatial-ICL-FT maintain relatively balanced results, further evidencing that the methods we propose are not dependent on the quantity of training samples but rather on the internal knowledge of the LLM and the quality of retrieved examples.

Additionally, this section compares the few-shot activation method with results from prior fully supervised models trained on full datasets, as shown in Table II. The data reveal that Spatial-ICL-FT improves the F1 score by 9.6 points over the R-BERT model, indicating that even in a few-shot scenario,

TABLE I: Experimental results comparing different methods in few-shot settings with various models

Model	QSLINK			OLINK			MOVELINK			Overall		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Supervised Learning (100 samples)												
R-BERT	35.2	38.5	36.8	61.0	55.2	60.0	51.7	53.2	52.4	50.8	60.9	55.4
HMCGR	43.2	56.1	48.8	52.0	62.2	56.6	50.8	65.2	57.1	50.0	64.1	56.2
AMR-Enhanced	61.7	66.2	63.9	50.6	55.7	53.0	66.3	62.9	64.6	59.2	61.6	60.4
In-Context Learning (12-Shot)												
ICL-Random	65.3	59.4	62.2	57.7	53.1	55.3	69.8	67.2	68.5	64.3	59.9	62.0
ICL-kNN	67.8	62.1	64.8	58.1	55.0	56.5	73.3	69.5	71.3	66.4	62.2	64.2
Spatial-ICL-SimCSE	72.1	67.3	69.6	68.5	69.2	68.8	71.8	68.3	70.0	70.8	68.3	69.5
Spatial-ICL-FT	**74.0**	**68.9**	**71.4**	**69.8**	**70.7**	**70.2**	**72.4**	**69.1**	**70.7**	**72.0**	**69.6**	**70.8**

TABLE II: Comparison results table with fully supervised models trained on full datasets

Model	P	R	F1
Spatial-ICL-FT	72.0	69.6	70.8
R-BERT	62.7	59.8	61.2
HMCGR	64.3	**79.2**	70.9
AMR-Enhanced	**75.0**	76.9	**76.0**

TABLE III: Comparison of Reasoning Enhancement Module

Methods	K=3	K=6	K=9	K=12
Spatial-ICL-SimCSE	57.3	59.8	65.9	69.5
w/o Reasoning	42.4	52.3	59.8	66.7
Spatial-ICL-FT	61.4	65.2	69.6	70.8
w/o Reasoning	44.2	53.8	60.5	67.8

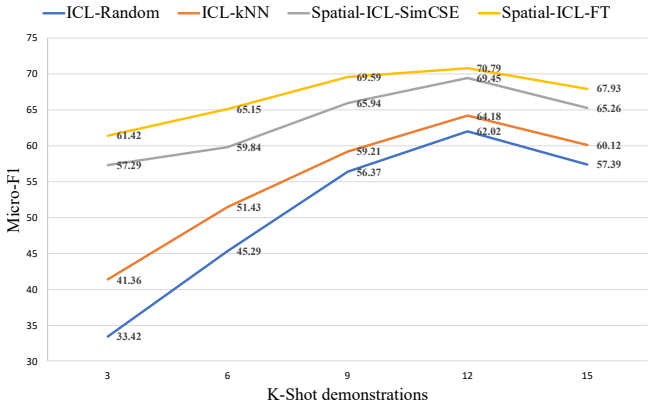

Fig. 3: Comparison of Task-Oriented Module Across Different K Values

our proposed model can outperform fully fine-tuned models trained on complete datasets. However, its F1 score is 5.2 points lower than that of the AMR-Enhanced model, suggesting that LLMs still lack the task-specific knowledge required for this domain and have not yet fully realized their potential in this task. Although large language models have achieved groundbreaking potential through ICL, their performance in spatial relation classification still lags behind state-of-the-art fully supervised models.

C. Ablation Analysis of Task-Oriented Module

This subsection first conducts an ablation analysis of the example retrieval module, evaluating its performance impact by incrementally increasing the number of K-shot examples, as shown in Figure 3.

The experimental results reveal that, compared to ICL-Random, all retrieval-based models exhibit higher F1 scores and more pronounced performance growth trends. This suggests that GPT-4 can more effectively learn classification patterns from high-quality spatial relation examples. After incorporating entity information into SimCSE retrieval, Spatial-ICL-SimCSE demonstrates superior performance across all K-shot settings, indicating that task-oriented sentence representations can more accurately capture the semantic features relevant to spatial relation classification, thereby providing examples better aligned with task requirements. Spatial-ICL-FT, which uses fine-tuned representations optimized for spatial relations, outperforms other retrieval methods. Notably, even at $k = 3$, Spatial-ICL-FT achieves an F1 score of 61.42, surpassing the 59.84 of Spatial-ICL-SimCSE at $k = 6$. This result underscores that, in few-shot settings, the quality of examples has a greater impact on performance than their quantity. The figure also shows that, to a certain extent, increasing the number of examples improves spatial relation classification performance; however, beyond a threshold, additional examples introduce noise, leading to a decline in classification effectiveness.

D. Ablation Analysis of Reasoning Enhancement Module

This subsection analyzes the impact of reasoning-enhanced examples on model performance, with results presented in Table III. We made trade-offs between adding reasoning information and increasing the number of examples. The experimental results reveal several insights: incorporating reasoning-enhanced examples consistently improves GPT-4's performance over unenhanced examples, a trend observed in both Spatial-ICL-SimCSE and Spatial-ICL-FT. This indicates that reasoning guidance based on gold spatial relation labels effectively enhances GPT-4's semantic reasoning capabilities,

further optimizing the effectiveness of in-context learning by deepening the understanding of examples. Additionally, after removing reasoning information from Spatial-ICL-SimCSE, its F1 score of 66.7 remains higher than ICL-kNN's 64.2, highlighting the effectiveness of the Spatial-ICL method in introducing sentence representations enhanced with spatial elements.

Moreover, for Spatial-ICL-FT, as the number of examples increases, the performance gains from reasoning enhancement gradually diminish. This phenomenon can be explained by the fact that, with more high-quality examples, GPT-4 can independently capture the spatial relation logic embedded in the examples, reducing the additional contribution of reasoning to a saturation point. The effect of reasoning enhancement is particularly pronounced in few-shot scenarios, making this method an effective strategy for low-resource spatial relation classification. The goal in low-resource settings is to accurately identify spatial relations with sparse or even unlabeled data, a direction that will be a key focus of future exploration.

E. Case Study

This section presents specific examples of spatial relation classification to illustrate the effectiveness of the Spatial-ICL method we propose. For the sentence "I want to show that there are people in the United States that care about this," which contains the triplet (*people*, *in*, *United States*), the relation is QSLINK. In traditional ICL-kNN, the retrieved example is "We elucidated how we reached the confluence point on a court from a homestead," with the triplet (*point*, *on*, *court*) and a spatial relation of OLINK. Although kNN retrieval based on sentence embeddings can retrieve examples semantically similar to the test sentence as a whole, it primarily reflects overall semantic relations, deviating from the focus of the spatial relation task. In contrast, the Spatial-ICL method improves retrieval specificity, retrieving the example "There are many places in the city where they don't connect," which contains the triplet (*places*, *in*, *city*) with a QSLINK spatial relation. The Spatial-ICL method retrieves ICL examples more closely aligned with the test sample, providing the LLM with more directly relevant information and thereby enhancing the stability of the model's classification.

V. Conclusion

We propose a spatial relation classification method based on few-shot in-context learning, which incorporates two strategies to optimize performance: task-oriented example retrieval and gold label guided reasoning, aiming to narrow the performance gap with fine-tuned baseline models. Experimental results demonstrate that our model outperforms the existing GPT-4-based ICL baselines. Our future work will focus on extending Spatial-ICL to more fine-grained spatial relation typologies, dynamic retrieval strategies, and efficient reasoning mechanisms to further enhance generalization and real-world applicability.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments on this paper. This research was supported by the National Natural Science Foundation of China (Nos. 62276177 and 62376181).

References

[1] X. Ma, "Knowledge graph construction and application in geosciences: A review," *Computers & Geosciences*, vol. 161, p. 105082, 2022.

[2] A. Mishra and S. K. Jain, "A survey on question answering systems with classification," *Journal of King Saud University-Computer and Information Sciences*, vol. 28, no. 3, pp. 345–361, 2016.

[3] B. J. Gutiérrez, N. McNeal, C. Washington, Y. Chen, L. Li, H. Sun, and Y. Su, "Thinking about gpt-3 in-context learning for biomedical ie? think again," in *Findings of EMNLP 2022*, 2022, pp. 4497–4512.

[4] K. Valmeekam, A. Olmo, S. Sreedharan, and S. Kambhampati, "Large language models still can't plan (a benchmark for llms on planning and reasoning about change)," in *Proceedings of NeurIPS 2022 Foundation Models for Decision Making Workshop*, 2022.

[5] J. Liu, D. Shen, Y. Zhang, W. B. Dolan, L. Carin, and W. Chen, "What makes good in-context examples for gpt-3?" in *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, 2022, pp. 100–114.

[6] Z. Wan, F. Cheng, Z. Mao, Q. Liu, H. Song, J. Li, and S. Kurohashi, "GPT-RE: In-context learning for relation extraction using large language models," in *Proceedings of EMNLP*, 2023, pp. 3534–3547.

[7] J. Pustejovsky, P. Kordjamshidi, M.-F. Moens, A. Levine, S. Dworman, and Z. Yocum, "Semeval-2015 task 8: Spaceeval," in *Proceedings of SemEval*, 2015, pp. 884–894.

[8] J. Pustejovsky, J. L. Moszkowicz, and M. Verhagen, "Using iso-space for annotating spatial information," in *Proceedings of the International Conference on Spatial Information Theory*, 2011.

[9] E. Nichols and F. Botros, "Sprl-cww: Spatial relation classification with independent multi-class models," in *Proceedings of SemEval*, 2015, pp. 895–901.

[10] J. D'Souza and V. Ng, "Sieve-based spatial relation extraction with expanding parse trees," in *Proceedings of EMNLP*, 2015, pp. 758–768.

[11] H. Salaberri, O. Arregi, and B. Zapirain, "Ixagroupehuspaceeval:(x-space) a wordnet-based approach towards the automatic recognition of spatial information following the iso-space annotation scheme," in *Proceedings of SemEval*, 2015, pp. 856–861.

[12] N. Ramrakhiyani, G. Palshikar, and V. Varma, "A simple neural approach to spatial role labelling," in *Proceedings of ECIR*, 2019, pp. 102–108.

[13] H. J. Shin, J. Y. Park, D. B. Yuk, and J. S. Lee, "Bert-based spatial information extraction," in *Proceedings of the Third International Workshop on Spatial Language Understanding*, 2020, pp. 10–17.

[14] S. Wu and Y. He, "Enriching pre-trained language model with entity information for relation classification," in *Proceedings of CIKM*, 2019, pp. 2361–2364.

[15] F. Wang, P. Li, and Q. Zhu, "A hybrid model of classification and generation for spatial relation extraction," in *Proceedings of COLING*, 2022, pp. 1915–1924.

[16] G. Yang, S. Xu, P. Li, and Q. Zhu, "Spatial relation extraction on AMR enhancement and additional markers," in *Proceedings of ICIC*, 2024, pp. 434–445.

[17] T. Gao, X. Yao, and D. Chen, "SimCSE: Simple contrastive learning of sentence embeddings," in *Proceedings of EMNLP*, 2021, pp. 6894–6910.

[18] Z. Zhong and D. Chen, "A frustratingly easy approach for entity and relation extraction," in *Proceedings of NAACL-HLT*, 2021, pp. 50–61.

[19] Z. Wan, Q. Liu, Z. Mao, F. Cheng, S. Kurohashi, and J. Li, "Rescue implicit and long-tail cases: Nearest neighbor relation extraction," in *Proceedings of EMNLP*, 2022.

[20] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," in *Proceedings of NeurIPS*, vol. 35, 2022, pp. 24 824–24 837.

[21] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," in *Proceedings of NeurIPS*, vol. 35, 2022, pp. 22 199–22 213.