Front. Comput. Sci., 2023, 0(0): 1–17
https://doi.org/10.1007/sxxxxx-yyy-zzzz-1

REVIEW ARTICLE

Integrating Element Correlation with Prompt-based Spatial Relation Extraction

Feng WANG[1], Sheng XU[1], Peifeng LI(✉)[1], Qiaoming ZHU[1]

1 School of Computer Science and Technology, Soochow University, Suzhou 215000, China

Abstract Spatial relations in text refer to how a geographical entity is located in space in relation to a reference entity. Extracting spatial relations from text is a fundamental task in natural language understanding. Previous studies have only focused on generic fine-tuning methods with additional classifiers, ignoring the importance of the semantic correlation between different spatial elements and the large offset between the relation extraction task and the pre-trained models. To address the above two issues, we propose a spatial relation extraction model based on Dual-view Prompt and Element Correlation (DPEC). Specifically, we first reformulate spatial relation extraction as a mask language model with a dual-view prompt (i.e., Link Prompt and Confidence Prompt). Link Prompt can not only guide the model to incorporate more contextual information related to the spatial relation extraction task, but also better adapt to the original pre-training task of the language models. Meanwhile, Confidence Prompt can measure the confidence of candidate triplets in Link Prompt and work as a supplement to identify those easily confused examples in Link Prompt. Moreover, we incorporate the element correlation to measure the consistency between different spatial elements, which is an effective cue for identifying the rationality of spatial relations. Experimental results on the popular SpaceEval show that our DPEC significantly outperforms the SOTA baselines.

Keywords Spatial Relation Extraction, Dual-view Prompt, Spatial Element Correlation, Link Prompt, Confidence Prompt

1 Introduction

Spatial relation extraction is to recognize the relation between two geographical entities in text. Most existing studies on relation extraction aim at extracting entity, temporal and causal relations, while only a few studies explore spatial relations. However, spatial information is a kind of critical information for natural language understanding

Received month dd, yyyy; accepted month dd, yyyy

E-mail: pfli@suda.edu.cn

and autopilot including spatial domain query [1], spatial reference [2] and data forecasting [3].

SpaceEval [4] proposed an annotation scheme adopted from ISO-space [5] for spatial relation extraction, which is widely used in previous work. In general, this task classifies the static and dynamic spatial relations into three types: MOVE LINK (MOVELINK), Qualitative Spatial LINK (QSLINK), and Orientation LINK (OLINK). MOVELINK connects motion-events with corresponding mover-participants as a triplet *(mover, goal, trigger)*, while QSLINK and OLINK refer to the topological and non-topological relation between two spatial elements, respectively, and a triplet is formalized as *(trajector, landmark, trigger)*.

Following previous work, we also simplify the whole spatial relation extraction task as a triplet extraction task, and the examples are shown in Figure 1. In general, spatial relations consist of spatial elements and spatial roles. Taking the QSLINK triplet (*woman*, *in*, *schoolyard*) in figure 1 as an example, QSLINK is the relation type, while "woman", "in", and "schoolyard" are three spatial elements that play the spatial roles of *trajector*, *trigger*, and *landmark* in the relation, respectively.

Sentences:

A [woman] is sitting [at] [table] [in] the middle of a [schoolyard] . After a while, [she] [walk out].

Relations:

{ MOVELINK	{ QSLINK	{ QSLINK
mover=she	trajector=woman	trajector=woman
trigger=walk out	trigger=at	trigger=in
goal=null }	landmark=table }	landmark=schoolyard}

Fig. 1 An example of spatial relation extraction where three spatial relations are extracted from the sentences.

In Fig. 1, spatial relation extraction first extracts three triplets (*she*, *walkout*, *null*[1)]), (*woman*, *at*, *table*) and (*woman*, *in*, *schoolyard*), and then assigns a role to each element in a triplet (e.g., the role of "table" is *landmark*). Finally, it assigns a spatial relation to each triplet (e.g., the spatial relation of (*woman*, *at*, *table*) is QSLINK).

Obviously, the elements among a triplet are interdependent. Taking the triplet "woman (is sitting) at table" in Figure 1 as an example, it is obviously unreasonable to replace "at" with "in" because a woman cannot be sitting in a table. Obviously, the trigger (e.g., "at" in the above triplet) is a very important evidence for extracting spatial relation. In a spatial relation, the trigger plays the core role in all elements and can act as a bridge to semantically connect the other two elements. Therefore, how to represent the trigger semantics in a spatial relation is the first challenge in spatial relation extraction.

In general, humans often guide spatial relations according to the semantic correlations between different elements. For example, when judging whether "woman (is sitting) at table" is a QSLINK relation, we first consider the semantic correlations between two elements "woman" and "table", and between the above two elements and the core element "at" which plays as a trigger role in the relation. Thus, how to capture the element correlations between spatial elements is the second challenge in spatial relation extraction.

Therefore, the trigger semantics and the element correlations between spatial elements are critical shreds of evidence for spatial relation extraction. However, previous studies only embedded those triplets using pre-trained models, ignoring the semantic relations among three spatial elements and the core role of the trigger. For example, a triplet (*table*, *walk out*, *schoolyard*) is invalid because a

1)Here *null* refers to a null element that cannot be extracted from the sentence.

table cannot walk normally. Hence, this triplet will be identified as NOLINK, because it does not belong to any predefined relation types. In this paper, we first introduce the element correlation to measure the consistency between different spatial elements.

The pre-trained model has become a hot research topic [6, 7] due to its superior performance on a wide range of language-related downstream tasks. Recent studies added additional classifiers on top of Pre-trained Language Models (PLMs) following the steps of pre-training and fine-tuning. For example, some studies on spatial relation extraction [8, 9] used a relation extractor on the top of BERT. Despite the success of fine-tuned PLMs, some recent studies [10, 11] find their challenges are the significant gap between objective forms in pre-training and fine-tuning, which restricts taking full advantage of PLMs. Inspired by prompt-tuning, we further reformulate the spatial relation extraction as a Mask Language Model (MLM) with multiple prompt views to bridge this gap.

In this paper, we propose a spatial relation extraction model with Dual-view Prompt and Element Correlation (DPEC). We first reformulate spatial relation extraction as a mask language model with a dual-view prompt (i.e., Link Promt and Confidence Prompt). Link Prompt focuses on the spatial element correlation, which can not only guide the model to incorporate more contextual information related to the spatial relation extraction task, but also better adapt to the original pre-training task of the language models. Meanwhile, Confidence Prompt aims at trigger semantics, which can measure the confidence of candidate triplets in Link Prompt and work as a supplement to identify those easily confused examples in Link Prompt. Moreover, we incorporate the element correlation to measure the consistency between different spatial elements, which is an effective cue to identify the rationality of spatial relations. Experimental results on the popular SpaceEval dataset show that our DPEC significantly outperforms the SOTA baselines. We will release the code at https://github.com/ZurichRain/DPED and our contributions are summarized as follows.

- We propose the Link Prompt and the element correlation mechanism to address the challenge of capturing the relations between spatial elements.
- We propose the Confidence Prompt to address the challenge of capturing trigger semantics, which can work as a supplement to identify those easily confused examples in Link Prompt.
- Experimental results on SpaceEval show that DPEC outperforms the SOTA baselines significantly.

2 Related work

In this section, we first introduce spatial relation extraction from two aspects, i.e., datasets and methods, and then review the prompt methods used in this paper.

2.1 Datasets

Previous studies designed various schemes to represent spatial relations. SpatialML [12] proposed the orientation and topological relations between locations according to the area calculus. SpatialML identified and marked locations mentioned in text using PLACE tags, and established the onnection between two locations. In addition to indicating the location itself, these tags can also include geographical coordinates recorded in a latLong (i.e., latitude and longitude) attribute. Orientation relations can further refine the location information, with the mod (i.e., the direction of north, south, east and west) attribute of the PLACE tag containing predefined placeholders to represent these orientations.This dataset has 947 documents and 15,004 PLACE tags.

SpRL [13] focused on the main roles in spatial relations and developed a semantic role labeling scheme. It utilized the standard test and training data provided by the SemEval-2007 challenge, which consisted of 34 XML files. Each file corresponded to a different preposition, and the dataset contained over 25,000 instances, with 16,557 examples in the training set and 8,096 in the test set. Each sentence in the dataset included a single occurrence of the respective preposition.

Spatial relation extraction is a subtask of SemEval 2012 [14], 2013 [15] and 2015 [4]. In SemEval 2012 [14] , it emphasized the primary roles of trajectors, landmarks, and spatial indicators, as well as the connections between these roles that give rise to spatial relations. The formal semantics of these relations was examined at a high-level, with three main types identified: directional, regional (topological), and distal. It contained mostly static spatial relations. In SemEval 2013 [15], this task was extended to include the identification of motion indicators and paths, which are relevant for more dynamic spatial relations.

SpaceEval in SemEval 2015 [4] proposed an annotation scheme adopted from ISO-space and enriched the semantics of SpRL by refining the granularity, while incorporating additional attributes of spatial elements and relations. Most previous studies were evaluated on this dataset and it is widely regarded as the most authoritative source of data for extracting spatial relations. The dataset defined three spatial relation types as follows:

QSLINK: which is used in ISO-Space to capture the topological relations between tagged elements (e.g., *in*, *connected* and *disconnected*).

OLINK: orientation links describe non-topological relations (e.g., *north*, *left*, *down* and *behind*).

MOVELINK: which is used as a dynamic spatial relation.

It also defined five roles as follows:

trajector: the entity, i.e., person, object or event whose location is described, which can be static or dynamic (also called local/figure object, locatum).

trigger: the core word to trigger a spatial relation, who is generally verbs or prepositions.

landmark: the reference entity in relation to which the location or the motion of the trajector is specified (also called reference object or relatum).

mover: things that act as movers in MOVELINK (e.g., person, object or event).

goal: it serves as a destination for movement in MOVELINK.

In the SemEval 2015 dataset, there are 59 annotated documents, with 1,110 QSLINKs, 287 OLINKs, and 974 MOVELINKs. Each document is saved in XML format.

2.2 Spatial relation extraction

Previous studies on spatial relation extraction can be divided into traditional machine learning and fine-tuned pre-trained methods.

2.2.1 Traditional machine learning

Traditional machine learning methods rely heavily on manual features or basic language structures. Nichols et al. [16] used a CRF layer to extract spatial elements and then introduced SVM to classify spatial relations. The system employs several techniques to extract meaningful elements from its input, including word embedding using GloVe [17], named entities, part-of-speech tags, and dependency parse labels. D'Souza et al. [18] proposed a Sieve-based model using various manual features through greedy feature selection techniques. They initially designed over 100 features and used a greedy approach to select the best features for jointly extracting spatial elements and spatial relations. Salaberri et al. [19] added external knowledge from WordNet and PropBank as a supplement to spatial information, which can better enrich the representation of spatial element.

Neural networks are also used to extract spatial relations. Ramrakhiyani et al. [8] used BiLSTM to classify the candidate relations generated by dependency parsing. Specifically, they propose a simple two-stage neural network approach for detecting static spatial relations.

2.2.2 Fine-tuned pre-training

Many studies preferred to utilize pre-trained and fine-tuned methods due to their excellent performance on downstream tasks. Shin et al. [9] first used BERT-CRF to extract the spatial roles and then introduced R-BERT [20] to extract the spatial relations. Wang et al. [21] proposed a hybrid model of generation and classification for this task, where the classifier can generate those non-null-role relations and the generator can extract those null-role relations to complement each other. The core practice of these methods is to add an adapter on the top of BERT, which does not take into account the offset between the spatial relation extraction task and the pre-training task.

2.3 Prompt-based methods

To further leverage knowledge embedded in pre-trained models, prompt tuning has been proposed to fill the gap between the training process in pre-training and fine-tuning. Shin et al. [10] proposed an automated method to create prompts for a diverse set of tasks, based on a gradient-guided search. Chia et al. [22] used prompting language models to generate structured texts in the zero-shot entity relation extraction. Han et al. [23] proposed prompt tuning with rules and applied logic rules to construct prompts with several sub-prompts for those complicated classification tasks. Different from the above work, we propose a dual-view prompt model, where Confidence Prompt can further assist in identifying those spatial triplets with low confidence in Link Prompt.

3 Spatial relation extraction model

Our DPEC consists of two components, i.e., candidate triplet extraction and spatial relation classification. As shown in Figure 2, the overall architecture of our model DPEC can be described as follows.

In the stage of candidate triplet extraction, we first introduce a BERT-CRF to identify spatial elements and then obtain the set of candidate triplets by arranging the spatial elements.

In the stage of spatial relation classification, we first generate two prompt templates, i.e., Link and Confidence Prompt templates, respectively, according to the set of candidate triplets, and then concatenate the original sequence text and two prompt templates as two input sequences to BERT. Thus, we obtain the representations of the [MASK] tokens in Link Prompt for spatial relation extraction and those in Confidence Prompt for trigger recognition. Finally, taking into account the inherent clustering of spatial elements in terms of semantic correlation between different elements, we fuse the representation of the element correlation

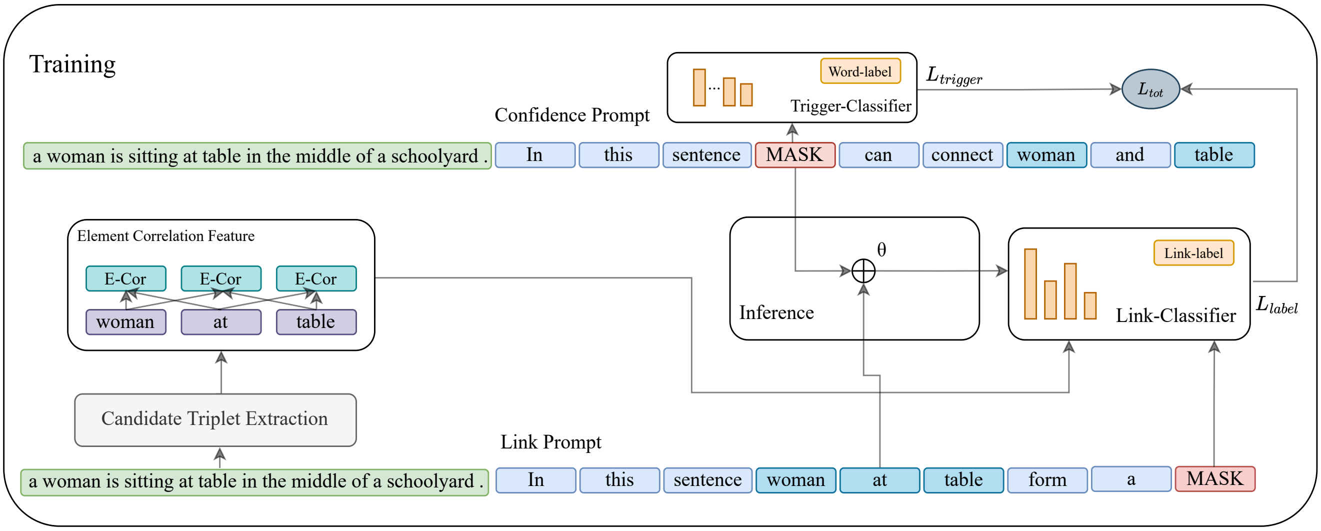

Fig. 2 Overall structure of our model DPEC.

between the spatial elements in the Link Prompt classifier. Moreover, we train the two tasks simultaneously in the training step and take the [MASK] result of Confidence Prompt as the evaluation for the Link Prompt classifier in the inference step.

3.1 Task definition

Our pipeline approach for the spatial relation extraction task can be divided into three subtasks, i.e., spatial element extraction, spatial role recognition, and spatial relation classification, where the task of candidate triplet extraction consists of the first two subtasks. Spatial elements are text fragments containing spatial information, while spatial roles are the roles played by spatial elements in spatial relations. For example, in the QSLINK relation shown in Figure 1, "woman" is a spatial element (spatial entity), and $trajector$ is the spatial role played by the woman.

The task of spatial element extraction is to extract possible elements from text and assign an element type to each element. Let $S = (s_1, ..., s_i, ...s_n)$ represent the set of text segments in sentences, where n is the number of the text segments. For each segment s_i in S, the goal of spatial element extraction is to learn a model $f_e : s_i \rightarrow y_e(s_i \in S)$ to classify each text segment into a predefined element type y_e ($y_e \in \{$ *Spatial Entity*, *Spatial Signal*, *Place*, *Motion Signal*, *Path*, *Motion*, *Non Motion Event*, *Measure*$\}$).

The task of spatial role recognition is to assign a role to each extracted element. Let $SE = \{se_1, ..., se_i, ..., se_m\}$ represent the set of the spatial elements extracted by spatial element extraction, where m is the number of the extracted elements. For each element se_i in SE, the goal of spatial role recognition is to learn a model $f_{role} : se_i \rightarrow y_{role}(se_i \in SE)$ to classify each element into a predefined role y_{role} ($y_{role} \in \{mover, trajacter, goal, landmark, trigger\}$).

According to the results of spatial element extraction, we can numerate all possible triplets as candidates according to the spatial relation definition. Let $CT = \{ct_1, ..., ct_i, ..., ct_k\}$ represent the set of all candidate triplets, where k is the number of the candidate triplets. For each ct_i in CT, the goal of spatial relation classification is to learn a model $f_{rel} : ct_i \rightarrow y_{rel}(ct_i \in CT)$ to classify each can-

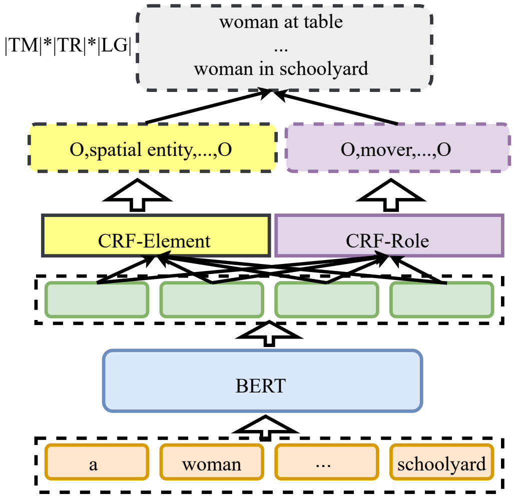

Fig. 3 Candidate triplet extraction.

didate triplet into a predefined relation y_{rel} ($y_{rel} \in \{MOVELINK, QSLINK, OLINK, NOLINK\}$) where $NOLINK$ refers to pseudo triplet.

3.2 Candidate triplet extraction

Candidate triplet extraction is to extract spatial relation triplets from text and feed them to spatial relation classification to identify their spatial relations, which includes two subtasks, i.e. spatial element extraction and spatial role recognition. Specifically, candidate triplet extraction is designed to extract what spatial element a text segment refers to and what spatial role it plays in order to better construct candidate spatial relation triplets.

Since each spatial relation is represented as a triplet with a corresponding relation type (MOVELINK, OLINK or QSLINK), the goal of our candidate triplet extraction is to extract as many candidate triplets as possible from raw text. Similar to HMCGR [21], we also use BERT+CRF for spatial role extraction, as shown in Figure 3, which first extracts the spatial elements from the text and then assigns a role to each extracted element. Formally, we simply take a text sequence $S = (s_1, ..., s_i, ..., s_n)$ as the input and then encode S as $H_B = [b_1, ..., b_i, ..., b_n]$ using BERT where $b_i \in R^{d_b}$ and R^{d_b} is the space of embeddings.

Since joint training of the two tasks can mutually improve the recognition performance, we apply a multitask framework to train spatial role recognition and spatial element extraction using two CRFs (i.e., CRF-Role and CRF-Element) on top of BERT, which share the same BERT encoder layer. We feed H_B to CRF-Role and CRF-Element, respectively to obtain the token labels. As for the output of the two CRF-based layers, we use O to represent non-element or non-role, and use element types and spatial roles (i.e., Y_e and Y_r in Subsection 3.1) to indicate specific elements in the spatial element extraction and spatial role recognition tasks, respectively. Taking the sentence in Figure 1 as an example, we can extract the spatial elements "woman", "table", "at", "in", "schoolyard" and "walk out", whose element types are *Spatial Entity*, *Spatial Entity*, *Spatial Signal*, *Spatial Signal*, *Place* and *Motion Signal*, respectively.

Since the performance of spatial role recognition is lower than that of spatial element extraction, we only use the results of the spatial element extraction as the candidate elements to construct spatial triplets. Thus, the spatial role recognition task is only an auxiliary task to enhance the spatial element extraction task.

In order to generate as many positive samples as possible for the next step spatial relation classification, we tend to generate all possible spatial role triplets from the set of the spatial elements. Specifically, we divide the spatial elements into three subsets, i.e., 1) TM={*trajector*, *mover*}, 2) LG={*landmark*, *goal*}, and 3) TR={*trigger*}, according to their element types. Taking the sentence in Figure 1 as example, “woman” and “table” belong to TM, “table” and“schoolyard” belong to LG, and “at”, “in” and “walk out” belong to TR.

Finally, we enumerate all possible triplets as candidates according to the spatial relation definition. As such, we extracted the candidate triplet set $ET = |TM| * |TR| * |LG|$ of the sentence in Figure 1 as follows: (*woman*, *at*, *table*), (*woman*, *in*, *schoolyard*), (*table*, *in*, *schoolyard*), (*table*, *at*, *schoolyard*), ... , and there are $12(2 \times 2 \times 3)$ candidate triplets in total.

3.3 Dual-view prompt

Inspired by the success of prompt-tuning in NLP, we reformulate spatial relation classification as a MLM task with dual-view prompt, i.e., Link and Confidence Prompt, from the viewpoints of spatial relation and spatial trigger, respectively. The Link Prompt is to predict spatial relations according to spatial element corrections, while the Confidence Prompt is to predict triggers based on the other two elements in the triplet. Adding Link Prompts to the original sentence can not only guide the model to incorporate more contextual information related to the task, but also better adapt to the original pre-training task of the language models. Moreover, we noticed that some triplets (e.g., (*woman*, *in*, *table*) in Figure 1) are indistinguishable using only the Link Prompt. Therefore, we propose another prompt, Confidence Prompt, to measure the confidence of candidate triplets in Link Prompt and work as a supplement to identify those easily confused examples in Link Prompt. If the prediction is correct, it will assign a high confidence to the Link Prompt, otherwise a lower one.

We first generate two prompt templates according to the set of candidate triplets extracted by candidate triplet extraction. Specifically, we need two template sentences with [MASK] as the input of Link Prompt and Confidence Prompt.

Link Prompt To make the spatial relation classification task better adapted to PLM and make use of MLM’s inherent knowledge, we propose the Link Prompt template. Link Prompt regards a triplet as a sentence, which can better capture the contextual information and capture element correlations. The prompt part can extract the key information in a sentence, which pays more attention to the information closely related to spatial relations. In addition, this prompt can filter the pseudo triplets by using contextual semantic errors. The Link Prompt template TL can be formalized as a relation prediction task as follows:

TL: In this sentence *tm lg tr* form a [MASK].

where (*tm*, *lg*, *tr*) is a candidate triplet ($tm \in TM$, $lg \in LG$ and $tr \in TR$). The purpose of adding “In this sentence” and “form” is to encourage the model to pay more attention to the context of the sentence and guide the language model to predict the tokens that best match the spatial relations. For example, given a candidate spatial relation triplet *(woman, at, table)*, TL can be generated as “In this sentence *woman at table* form a *[MASK]*”.

Confidence Prompt Since parts of QSLINKs and OLINKs are difficult to distinguish, and some pseudo triplets may be assigned a spatial relation only using Link Prompt, we introduce Confidence Prompt to determine the confidence of a spatial triplet in the inference step. Confidence Prompt can not only assist Link Prompt in distinguishing between QSLINKs and OLINKs, but can also filter out these pseudo triplets. The Confidence Prompt template TC can be formalized as a trigger prediction task to discover trigger semantics as follows:

TC: [MASK] can connect *tm* and *lg*.

where the spatial relation is triggered by the word "connect", which makes the trigger recognition task more reasonable. For example, given a candidate triplet *(woman, at, table)*, TC can be generated as "*[MASK]* can connect *woman* and *table*".

After constructing the above two prompts, for each candidate triplet we simply concatenate TL/TC and their original sentences S as an input sequence which can be formulated as follows:

$$XL = concat(S, TL) \tag{1}$$

$$XC = concat(S, TC) \tag{2}$$

where $XL = \{xl_1, ..., xl_i, ..., xl_n\}$ and $XC = \{xc_1, ..., xc_i, ..., xc_m\}$. xl_i and xc_i are the i-th token in XL and XC, respectively.

Since feeding the prompt into the MLM can cast our task as a language modeling task, we feed XL and XC into the attentive MLM encoder (BERT) to obtain the argument interaction representation as follows:

$$HL = MLM\text{−}Encoder(XL) \tag{3}$$

$$HC = MLM\text{−}Encoder(XC) \tag{4}$$

3.4 Element correlation

For any spatial relation triplet, there must be some kind of semantic correlation between three elements. For example, since a table cannot walk, the triplet (*table*, *walk out*, *schoolyard*) does not exist spatial relation, as we mentioned above. Hence, humans can determine whether a triplet exists a certain spatial relation according to element correlations.

To simplify the evaluation method of element correlation in a triplet, a triple relation can be decomposed into three one-to-one sub-relations. For example in a MOVELINK triplet, we can evaluate the correlation of (*mover*, *trigger*), (*mover*, *goal*) and (*goal*, *trigger*), respectively. We found that the triggers of the dynamic and static relations are generally verbs and prepositions, respectively, while the affinity of trigger and trajector/mover/landmark/goal has different POSs (Part-Of-Speeches) in different relations. To take advantage of this regular pattern, we propose the element correlation E-Cor for (TM, TR), (TM, LG) and (TR, LG).

In addition to capturing the characteristics of sub-relations, E-Cor can also use the trigger's POS and its value range to help identify spatial relations. Specifically, we first introduce the SelfAttentiveSpanExtractor in the NLP tool AllenNLP[2] to obtain the latent representations of three spatial roles H_{tm}, H_{lg} and H_{tr} as follows:

$$H_y = \sum_{i=y_{start}}^{y_{end}} (W_{y_i} hl_{y_i}) \tag{5}$$

where $y \in \{tm, lg, tr\}$. y_{start} and y_{end} represent the start and end position of a spatial element, respectively, hl_{y_i} is the representation of the i-th token in HL_y, and W_{y_i} are learnable parameters. Besides, since BERT maybe splits a word into multiple word-pieces, we also use SelfAttentiveSpanExtractor to obtain word-level representation.

With the latent representation of three spatial elements, we express three E-Cors (i.e., *tm*-*lg*, *tr*-*tm* and *lg*-*tr*) as follows, where $x_i, y_j \in \{tm, lg, tr\}$.

$$E_{(x_i, y_j)} = abs(H_{x_i} - H_{y_j}) \tag{6}$$

3.5 Classification

In the link classification, we can simply splice E-Cor and the mask token of the link prompt as the latent representation of the spatial relation, then we

[2] https://allenai.org/allennlp

use the softmax function σ to get the relation label of the triplet y_{rel} as follows.

$$y_{rel} = \sigma([HL_m; E_{(tm,lg)}; E_{(lg,tr)}; E_{(tr,tm)}]) \quad (7)$$

where HL_m is the mask representation of HL, and $[A;B]$ represents the concatenation of A and B.

For trigger classification, we assume that E-Cor carries trigger information. To let the model learn the trigger information by itself, and to give better confidence to the link prompt, we use only the mask as its latent representation. Then we also use softmax to get the trigger label y_{tri} as follows.

$$y_{tri} = \sigma(HC_m) \quad (8)$$

where HC_m is the mask representation of HC.

The two classifiers are activated using signal Softmax Layer and both of them use the Cross Entropy loss. Finally, we train the two tasks simultaneously by simply adding the loss as follows.

$$L_{tot} = L_{rel} + L_{tri} \quad (9)$$

where L_{rel} and L_{tri} are the losses of the Link classifier and the Confidence classifier, respectively.

3.6 Inference

In the inference step, we use only a simple rule to infer spatial relations. First, we compare the prediction results of the link classifier and the trigger classifier. If the predicted trigger is consistent with the trigger in the candidate triplet, this candidate triplet can form a spatial relation with high confidence; otherwise, the highest probability in the prediction result of the link classifier is set to 0.5 times of itself. The confidence of 0.5 provides the largest information entropy, and our experiments show that 0.5 is also the best on the development set.

Tool/Parameter	**Version/Value**
Pytorch	1.7.0+cu110
BERT	bert-base-uncased
Allennlp	2.6.0
Learning rate	2e-5
Batch size	8
Random seed	1024
Hidden size	768
Optimizer	AdamW
GPU	NVIDIA RTX 3090
CPU	i7

Table 1 Key parameters and tools used in our model.

Model	P	R	F1
BERT+CRF	88.1	91.2	89.1

Table 2 The results of spatial role extraction.

Taking the candidate triplet (*woman*, *at*, *table*) as an example, the link classifier identifies QSLINK as its spatial relation and the trigger classifier identifies “at” as its trigger. Since the predicted trigger “at” matches the one in the triplet, we assign QSLINK directly to this triplet. On the contrary, taking the pseudo triplet (*woman*, *in*, *table*) as an example, the probability distribution of the link classifier is [0.4, 0.2, 0.1, 0.3], which refers to (QSLINK, OLINK, MOVELINK, NOLINK), while the trigger predictor identifies “at” as its trigger. Since the predicted trigger “at” is inconsistent with the one (“in”) in the triplet, we replace the highest value 0.4 with 0.2 ([0.2, 0.2, 0.1, 0.3]) and then set its relation to NOLINK (0.3), i.e., pseudo triplet.

4 Experimentation

In this section, we first introduce the experimental settings and then report the experimental results. Finally, we analyze our model from different viewpoints.

Model	QSLINK			OLINK			MOVELINK			Overall		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
SpRL-CWW	66.1	53.8	59.4	69.1	51.7	59.1	57.1	45.1	50.4	63.6	50.1	56.1
R-BERT	45.1	58.3	50.5	71.0	69.6	70.2	62.7	61.5	62.1	62.7	59.8	61.2
HMCGR	53.5	73.1	61.1	**73.1**	**85.2**	**78.6**	66.8	83.0	73.9	64.3	79.2	70.9
DPEC	**81.0**	**80.8**	**80.9**	57.0	73.8	63.3	**91.3**	**92.0**	**91.6**	**76.4**	**82.2**	**79.2**

Table 3 Performance comparison between the baselines and our DPEC on SpaceEval. Since R-BERT did not report the results on each category, we run their sharing codes to obtain the results (underlined).

4.1 Experimental settings

We evaluate our model DPEC on the latest spatial dataset SpaceEval [4]. According to the official statistics, there are 1110 QSLINKs, 974 MOVELINKs, and 287 OLINKs. We use the standard training/development/test set following previous work [9] where the rate is 6:2:2. As for evaluation, we report Precision (P), Recall (R), and Micro-F1 score following previous work [21]. Furthermore, our model is implemented based on Pytorch and Huggingface. The specific tool versions and key hyper-parameters are showed in Table 1.

Currently, only a few work focused on spatial relation extraction. To evaluate the effectiveness of our model, we conduct the following three strong baselines for comparison:

- **SpRL-CWW** [16], which used SVM and CRF classifiers on the GloVe features to extract the spatial relations;
- **R-BERT** [9], which used a BERT-based neural network model and added an adapter on the top layer fine-tuning;
- **HMCGR** [21], which is the SOTA model using a hybrid model on the spatial relation extraction. They also combined classification and generation to extract those non-null-role relations.

4.2 Experimental results

This paper does not focus on spatial role extraction and its results on SpaceEval are shown in Table 2. Its performance is similar to those in previous work [9,21]. Due to the low recall in parts of previous work, we enumerate all possible triplets as candidates following HMGCR [21] to extract as many triplets as possible. Finally, we obtain 1185 candidate triplets from the test set, of which 474 triplets are positive and 711 are negative. Since more than half of the triplets are pseudo ones, we introduce NOLINK to represent these pseudo triplets.

Table 3 shows the overall performance of the three baselines and our model DPEC on the SpaceEval dataset. We used the t-test with a 95% confidence interval for the significance test and all improvements of DPEC over HMGCR and R-BERT are significant ($p<0.01$). We can find out that the three neural network models R-BERT, HMCGR, and DPEC outperform the other traditional SOTA model SpRL-CWW significantly. This result indicates the success of neural networks in the task of spatial relation extraction. Our model DPEC outperforms the SOTA model HMCGR rapidly, with gains of 8.3, 12.1, and 3.0 on F1-score, precision, and recall, respectively. This result shows the effectiveness of our DPEC in extracting spatial relations.

As for the different spatial relations, our DPEC

Model	QSLINK			OLINK			MOVELINK			Overall		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
DPEC	81.0	80.8	80.9	57.0	73.8	63.3	91.3	92.0	91.6	76.4	82.2	79.2
w/o Element Correlation	-18.7	-9.8	-15.4	-5.8	-13.1	-8.3	-22.3	-5.0	-15.6	-15.6	-9.3	-13.7
w/o Link Prompt	-10.3	-1.3	-6.2	-8.0	-10.1	-8.0	-8.6	-2.3	-5.9	-8.9	-4.5	-7.3
w/o Confidence Prompt	-2.0	0.0	-1.1	-2.0	0.0	-0.5	+1.0	0.0	+0.7	-1.0	0.0	-0.9
w/o L&C	-23.7	-5.5	-12.9	-4.5	-25.5	-16.8	-12.6	-3.0	-8.6	-13.6	-7.7	-12.4

Table 4 The results of our DEPD and its simplified versions on SpaceEval where L&C refers to Link Prompt and Confidence Prompt.

performs well on MOVELINK and QSLINK. However, it achieves a relatively low score on OLINK due to its small size in the training set. MOVELINK achieves the highest performance among the three relations due to its triggers, which are main verbs and different from the other two relations.

Compared with HMCGR, our DPEC greatly improves the F1 score on MOVELINK and QSLINK by 17.7 and 19.8, respectively, and the reason is that we introduce dual-view prompt and element correlation to identify the spatial relation and filter out those pseudo triplets. These results verify the effectiveness of the Link Prompt and element correlation in mining the relations among three spatial elements. However, the performance of OLINK of our DPEC is lower than that of HMCGR, and the reason is that HMCGR combined a classifier and a generator, in which the generator prefers to identify OLINK. In addition, a disadvantage of our DPEC is that it will assign pseudo-triplets to OLINK and harm the precision.

We analyze the confusion matrix of DPEC and find that there are only a few incorrectly predicted samples between MOVELINK and QSLINK/OLINK. However, most QSLINK triplets are misclassified as OLINK, or most OLINK triplets are misclassified as QSLINK. The reason is that the composition of QSLINK and OLINK is similar, which makes them difficult to distinguish, even for humans.

4.3 Analysis on element correlation

To further verify the effectiveness of our DPEC boosting on each module, we conduct the ablation experiments and Table 4 shows the results of the simplified versions. The performance decrease is obvious when we remove the element correlation (w/o Element Correlation). Compared with our final model DPEC, the F1-score drops by 13.7, indicating the effectiveness of the element correlation. We find that the decrease in precision is greater than the decrease in recall for each relation. This result indicates that the element correlation can not only discriminate the three relations, but also filter out the pseudo triplets to improve the precision.

We visualize the dimensionality reduction of the element correlation information as Figure 4 and the distribution of the triggers as Figure 5. Using the untrained BERT (upper line), we find that the distance between the elements *trajector/mover* (TM) and *trigger* (TR) and the distance between *landmark/goal* (LG) and *trigger* (TR) are somewhat distinguishable on QSLINK (blue) and MOVELINK (red). This result shows that the POS feature used in BERT is very helpful for identifying spatial relations. Using the element correlation in our DPEC (bottom line), the model can distinguish QSLINK

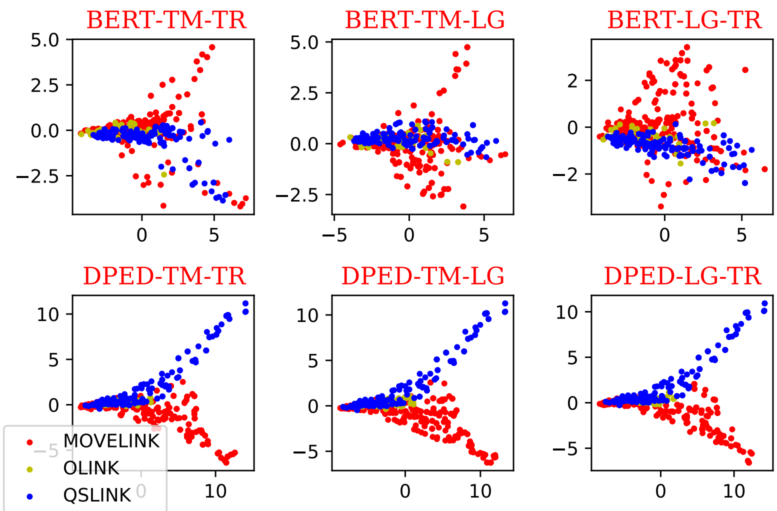

Fig. 4 Three sub-relation distances of spatial elements for BERT (upper line) and DPEC (bottom line), where the red, yellow and blue color refer to MOVELINK, OLINK and QSLINK, respectively.

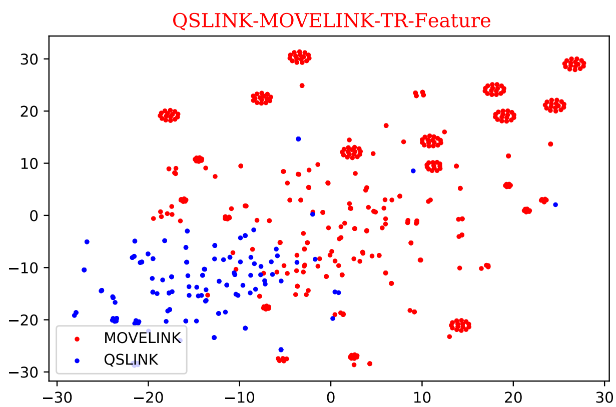

Fig. 5 Trigger distribution of QSLINKs and MOVELINKs.

Model	P	R	F1
DPEC	**76.4**	**82.2**	**79.2**
TM-TR	73.5	80.7	76.3
LG-TR	70.0	78.1	73.3
TM-LG	60.8	72.9	65.5

Table 5 Results on different sub-relations, where TM-TR refers to our model only using the distance of *trigger* and *landmark/goal*, TM-LG refers to our model only using the distance of *trajector/mover* and *landmark/goal*, and LG-TR refers to our model only using the distance of *trajector/mover* and *trigger*.

and MOVELINK easily, following all three distances. These trends also ensure the effectiveness of our element correlation mechanism on spatial relation extraction.

Figure 5 shows that the triggers are clustered in the semantic feature space and indicates that the triggers are also critical evidence for this task. Formally, the red circles are similar verbs, while some blue dots appear in the red area to indicate that some prepositions can be used as verbs (e.g., “toward” in “tail toward him”).

Table 5 shows the results of different sub-relations. TM-TR and LG-TR outperform TM-LG significantly and this indicates that the trigger is a critical cue to extract spatial relations, which is also verified in Figure 5. Moreover, TM-TR is better than LG-TR, and this result shows that the trajector and mover are more important than the landmark and goal in spatial relation extraction.

Table 6 shows the results of different methods to evaluate the element correlation. The bitwise multiplication (*) has the worst performance, while the bitwise subtraction (-) and cosine (cos) can achieve comparable performance compared to our absolute difference method used in DPEC. These re-

Model	P	R	F1
DPEC	**76.4**	**82.2**	**79.2**
*	70.5	75.7	73.0
-	75.2	81.9	78.4
cos	72.5	82.2	78.7

Table 6 Results on different methods of integrating element correlation, where "*", "-", and "cos" represents the bitwise multiplication, the bitwise subtraction and the cosine distance of two vectors.

Model	car on hill	we at lawn	woman in table
R-BERT	OLINK	NOLINK	QSLINK
LP	OLINK	QSLINK	QSLINK
CP	QSLINK	NOLINK	OLINK
L&C	QSLINK	QSLINK	NOLINK

Table 7 Examples of prediction in the simplified models of DPEC, where LP, LC and L&C refer to Link Prompt, Confidence Prompt and Dual-view Prompt.

sults suggest that using the absolute value difference method is the most effective way to measure the correlation between two spatial elements.

4.4 Analysis on dual-view prompt

Table 4 also shows the results of our two prompts and their combination. When we remove the Link Prompt (w/o Link Prompt, where TL is replaced by "[MASK]"), the F1-score drops by 7.3. This result shows that there is still a large offset between using vanilla pre-trained fine-tuning and the pre-trained BERT, and our Link Prompt is efficient in this task. Furthermore, using the simple feature concatenation at the top of BERT does not fully utilize the pre-trained model. On the contrary, by using a prompt to transform relation extraction into an MLM task, the model can more easily study the context of the current relation. Thus, Link Prompt can extract the key information in a sentence, which can pay more attention to the information closely related to the spatial relations. In addition, Link Prompt can also filter out pseudo triplets by using the contextual semantics, and the large drop in precision verifies this fact.

When we remove the Confidence Prompt (w/o Confidence Prompt), the F1 score drops by 0.9, mainly due to the precision, while the recall remains unchanged. This indicates that the Confidence Prompt is effective in filtering out these pseudo triplets and then improving Precision.

Although the improvement of the Confidence Prompt is relatively small, especially the small decrease of MOVELINK, it is a significant complement to the Link Prompt. Table 4 shows that the dual-view prompt can improve the overall F1 score, precision, and recall by 12.4, 13.6, and 7.1, respectively (DPEC vs. w/o L&C), which are larger than the sum of the improvements of the single Link and Confidence Prompts (P/R/F1: +3.7/+3.2/+4.2). This suggests that the two prompt tasks can complement each other when trained together.

Table 7 shows three examples of candidate triplets and their predicted relations by R-BERT, Link Prompt, Confidence Prompt, and Dual-view Prompt. R-BERT cannot recognize all three relations, while our L&C can recognize all three relations. For example, since "car on top of hill" is an OLINK, Link Prompt incorrectly identifies "car on hill" as an OLINK relation. On the contrary, Confidence Prompt can correctly assign QSLINK to this triplet because it can predict the correct trigger "on" and "car on hill" is a QSLINK. Also, the two prompt tasks can complement each other. Both Link and Confidence Prompt cannot correctly recognize the relation "woman in table", because the token "in" often triggers the OLINK and QSLINK relations. As mentioned above, since Confidence Prompt identifies "at" as its trigger and then reduces the probability of QSLINK according to our

Model	5%	30%	50%	75%	100%
DPEC	**50.4**	**59.2**	**70.2**	**73.3**	**79.2**
BERT	23.5	30.7	50.0	60.1	61.2
HMCGR	30.0	34.1	55.2	65.2	70.9

Table 8 F1 results on few-shot settings, where 5%, 30%, 50%, 75% and 100% represent the percentage of all training dataset.

inference mechanism, our L&C can correctly identify this relation.

4.5 Few-shot analysis

The existing work has shown that the use of prompt-based methods can achieve good results in few-shot learning tasks. In light of this, we conducted experiments with few-shot settings, where we divided the data into 5%, 30%, 50%, 75%, and 100% of the training data. We compared our results to strong baselines BERT and HMCGR, and the results are shown in Table 8. In Table 8, we can see that DPEC has smaller decreases than those of BERT and HMCGR in the above few-shot settings. This shows that our DPEC can achieve better results even in the case of few shots.

4.6 Case study

Example 1. Taking the sentence shown in Figure 1 as an example, both BERT and HMCGR identify (*she*, *walk out*, *schoolyard*) as a MOVELINK relation. However, it is obvious that in this sentence the goal is a missing value and "schoolyard" does not refer to the goal of the MOVELINK. Therefore, "walk out" cannot connect "she" and "schoolyard." Our DPEC identifies this triplet as NOLINK because the trigger prediction the value in the Confidence Prompt is "in," which does not match "walk out". After adjusting the probability distribution of the Link Prompt, the output link is calibrated to NOLINK. This indicates that the confidence prompt plays a crucial role in ensuring the accuracy of the model's predictions.

Example 2. In the sentence "We can see some flowers on the land, and there is an old tree whose roots are buried under the earth", the triplet (*roots*, *under*, *earth*) was identified as OLINK by both BERT and HMCGR. However, our DPEC can identify it as NOLINK correctly. This discrepancy can be explained by the fact that the phrase "root under" is used more often than "tree under" to describe spatial relations, and thus the distance between them is closer to the average distance between the trajector and the trigger in OLINK (TM-TR). Conversely, the semantic distance between "tree under" does not exhibit clustering properties, and therefore it is classified as NOLINK. This highlights the importance of incorporating prior knowledge in the form of element correlation into the process of determining spatial relations.

4.7 Error analysis

The majority of errors are due to these pseudo triplets. As mentioned above, in order to extract more candidate triplets, our candidate triplet extraction extracts 1185 candidate triplets, of which 711 (60.0%) triplets are pseudo triplets. These pseudo triplets affect the precision, because 7% of the pseudo triplets are misclassified as the three predefined relations, and 20% of the positive triplets are misclassified as NOLINK. In particular, almost all misclassified MOVELINKs are identified as NOLINK.

Since QSLINK and OLINK are very similar and their triggers are all prepositions, our model DPEC cannot effectively distinguish between these two relations. Taking the triplet (*palettes*, *on*, *heap*) in Figure 6 as an example, the implicit meaning of

Sentence: On the one hand I never saw more palettes on a heap before.

Gold QSLINK:	**Predicted OLINK:**
Trajector: palettes	Trajector: palettes
Trigger: on	Trigger: on
Landmark: heap	Landmark: heap

Fig. 6 Example of the errors in QSLINK and OLINK.

“on” as a trigger is with orientation, our model will misjudge it as OLINK.

We also find human-annotated labeling errors and unlabeled spatial relations in the dataset. For example, some spatial relations are transitive, such as “in” and “in front of”. However, some instances do not satisfy transitivity.

5 Conclusion

In this paper, we proposed a novel spatial relation extraction model on dual-view prompt and element correlation. Specifically, we reformulate spatial relation extraction as a dual-view prompt MLM task and introduce Link Prompt and Confidence Prompt to recognize spatial relations from the relation and trigger viewpoints, respectively. Moreover, we incorporate the element correlation to measure the consistency between different spatial elements, which are helpful to filter out pseudo samples. Experimental results on SpaceEval show that our proposed model DPEC outperforms the SOTA baselines significantly. Our future work will focus on the performance of OLINK and fine-grained spatial relation extraction.

Acknowledgements This work is supported by the National Natural Science Foundation of China (Nos. 62276177 and 61836007)) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

1. Zhang L, Wang R, Zhou J, Yu J, Ling Z, Xiong H. Joint Intent Detection and Entity Linking on Spatial Domain Queries. In: Findings of EMNLP 2020. 2020, 4937–4947
2. Yang T Y, Lan A S, Narasimhan K. Robust and Interpretable Grounding of Spatial References with Relation Networks. arXiv preprint arXiv:2005.00696, 2020
3. Song C, Lin Y, Guo S, Wan H. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In: AAAI. 2020, 914–921
4. Pustejovsky J, Kordjamshidi P, Moens M, Levine A, Dworman S, Yocum Z. SemEval-2015 Task 8: SpaceEval. In: Cer D M, Jurgens D, Nakov P, Zesch T, eds, SemEval. 2015, 884–894
5. Pustejovsky J, Moszkowicz J L, Verhagen M. Using ISO-Space for annotating spatial information. In: Proceedings of the International Conference on Spatial Information Theory. 2011
6. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Burstein J, Doran C, Solorio T, eds, NAACL-HLT. 2019, 4171–4186
7. Brown T, Mann B, Ryder N, Subbiah M, Kaplan J D, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, others . Language models are few-shot learners. NIPS, 2020, 33: 1877–1901
8. Ramrakhiyani N, Palshikar G K, Varma V. A Simple Neural Approach to Spatial Role Labelling. In: Azzopardi L, Stein B, Fuhr N, Mayr P, Hauff C, Hiemstra D, eds, ECIR. 2019, 102–108
9. Shin H J, Park J Y, Yuk D B, Lee J S. BERT-based Spatial Information Extraction. In: SPLU. November 2020, 10–17
10. Shin T, Razeghi Y, IV R L L, Wallace E, Singh S. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. In: Webber B, Cohn T, He Y, Liu Y, eds, EMNLP. 2020, 4222–4235
11. Schick T, Schütze H. It's not just size that matters: Small language models are also few-shot learners. In: Toutanova K, Rumshisky A, Zettlemoyer L, Hakkani-Tür D, Beltagy I, Bethard S, Cotterell R, Chakraborty T, Zhou Y, eds, NAACL-HLT. 2021, 2339–2352
12. Mani I, Doran C, Harris D, Hitzeman J, Quimby R, Richer J, Wellner B, Mardis S, Clancy S. SpatialML: annotation scheme, resources, and evaluation. Language Resources and Evaluation, 2010, 44(3): 263–280
13. Kordjamshidi P, Van Otterlo M, Moens M F. Spatial role labeling: Towards extraction of spatial relations from natural language. ACM Transactions on Speech

and Language Processing, 2011, 8(3): Article No. 4:1–36

14. Kordjamshidi P, Bethard S, Moens M F. SemEval-2012 task 3: Spatial role labeling. In: SemEval. 2012, 365–373
15. Kolomiyets O, Kordjamshidi P, Moens M F, Bethard S. Semeval-2013 task 3: Spatial role labeling. In: SemEval. 2013, 255–262
16. Nichols E, Botros F. SpRL-CWW: Spatial Relation Classification with Independent Multi-class Models. In: Cer D M, Jurgens D, Nakov P, Zesch T, eds, SemEval. 2015, 895–901
17. Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014, 1532–1543
18. D'Souza J, Ng V. Sieve-Based Spatial Relation Extraction with Expanding Parse Trees. In: Màrquez L, Callison-Burch C, Su J, Pighin D, Marton Y, eds, EMNLP. 2015, 758–768
19. Salaberri H, Arregi O, Zapirain B. IXAGroupEHUSpaceEval: (X-Space) A WordNet-based approach towards the Automatic Recognition of Spatial Information following the ISO-Space Annotation Scheme. In: Cer D M, Jurgens D, Nakov P, Zesch T, eds, SemEval. 2015, 856–861
20. Wu S, He Y. Enriching pre-trained language model with entity information for relation classification. In: CIKM. 2019, 2361–2364
21. Wang F, Li P, Zhu Q. A hybrid model of classification and generation for spatial relation extraction. In: COLING. 2022, 1915–1924
22. Chia Y K, Bing L, Poria S, Si L. Relationprompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction. In: Muresan S, Nakov P, Villavicencio A, eds, Findings of ACL. 2022, 45–57
23. Han X, Zhao W, Ding N, Liu Z, Sun M. PTR: prompt tuning with rules for text classification. CoRR, 2021, abs/2105.11259

Feng Wang received the B.S. degree from Shanxi University, China, in 2021. He is now a M.S. student in the School of Computer Science and Technology at Soochow University. His research interest lies in information extraction.

Sheng Xu received the M.S. degree from Soochow University, China, in 2019. He is now a Ph.D. student in the School of Computer Science and Technology at Soochow University. His research interest lies in discourse relation recognition and event relation extraction.

Peifeng Li received his Ph.D. degree in Computer Science from Soochow University in 2006. He has been a Professor in the School of Computer Science and Technology at Soochow University since 2015. His research interests include Chinese Computing, information extraction, etc.

Qiaoming Zhu received his Ph.D. degree in Computer Science from Soochow University in 2006. He is now a Professor in the School of Computer Science and Technology at Soochow University. His research interests include Chinese computing, discourse analysis, etc.