Prompt-based Zero-shot Relation Extraction with Semantic Knowledge Augmentation

Jiaying Gong, Hoda Eldardiry
Virginia Tech
Blacksburg, VA, USA
{gjiaying, hdardiry}@vt.edu

Abstract

In relation triplet extraction (RTE), recognizing unseen relations for which there are no training instances is a challenging task. Efforts have been made to recognize unseen relations based on question-answering models or relation descriptions. However, these approaches miss the semantic information about connections between seen and unseen relations. In this paper, We propose a prompt-based model with semantic knowledge augmentation (ZS-SKA) to recognize unseen relations under the zero-shot setting. We present a new word-level analogy-based sentence translation rule and generate augmented instances with unseen relations from instances with seen relations using that new rule. We design prompts with weighted virtual label construction based on an external knowledge graph to integrate semantic knowledge information learned from seen relations. Instead of using the actual label sets in the prompt template, we construct weighted virtual label words. We learn the representations of both seen and unseen relations with augmented instances and prompts. We then calculate the distance between the generated representations using prototypical networks to predict unseen relations. Extensive experiments conducted on three public datasets FewRel, Wiki-ZSL, and NYT, show that ZS-SKA outperforms other methods under zero-shot setting. Results also demonstrate the effectiveness and robustness of ZS-SKA.

Keywords: relation triplet extraction, zero-shot learning, knowledge augmentation

1. Introduction

Relation triplet extraction (RTE) aims to extract both the pairs of entities and relations from unstructured text. However, existing approaches based on supervised learning (Ma et al., 2022; Zhong and Chen, 2021; Ren et al., 2021; Huguet Cabot and Navigli, 2021; Zhu et al., 2019; Li and Tian, 2020) or few-shot learning (Liu et al., 2022b; Dou et al., 2022; Han et al., 2021a; Gao et al., 2019; Ren et al., 2020; Dong et al., 2020) still require labeled data. They can not catch up with a dynamic and open environment where new classes emerge. In the real-world setting, the classes of instances are sometimes rare or never seen in the training data. Thus, we tend to learn a model similar to the way humans learn and recognize new concepts. Such a task is referred to as zero-shot learning (ZSL). We follow the same definition of ZSL in (Chia et al., 2022; Chen and Li, 2021) to conduct experiments.

Zero-shot RTE aims to extract relation triplets in a sentence that is absent from the learning stage. Figure 1 shows an example of zero-shot RTE. The relation sets at the training and testing stages are disjoint. The model for zero-shot RTE is only trained on the seen relations in the training stage and extracts triplets with unseen relations in the testing stage. Existing approaches to zero-shot relation extraction still have limitations. First, some models perform zero-shot relation extraction by question answering (Levy et al., 2017) or by using GPT-2 to help generate synthetic data (Chia et al.,

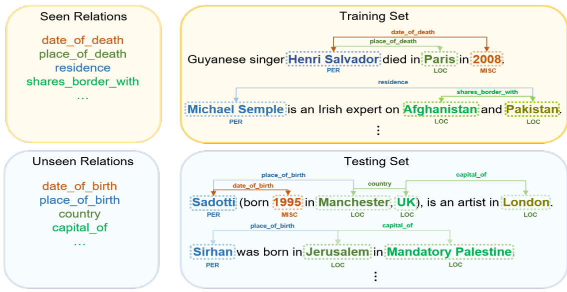

Figure 1: Zero-shot RTE. There is no overlap of classes between training and testing data.

2022). These models have a strong assumption that an excellent additional deep learning model is learned, and that all values extracted from this model are correct. Second, some existing studies formulate relation extraction as a text entailment task (Obamuyide and Vlachos, 2018). They only predict a binary label indicating whether the name entities in the given sentence can be described by a given description. Third, some SOTA models leverage auxiliary information to tackle zero-shot tasks. They focus on class names/descriptions, losing the connection or relationships between seen relations and unseen relations (Gong and Eldardiry, 2021; Chen and Li, 2021; Wang et al., 2022). Besides, these works mainly focus on zero-shot relation classification (ZSRC), which only predicts unseen relations instead of triplets in the format of <head entity, relation, tail entity>. ZSRC has a

LREC-COLING 2024, pages 13143–13156
20-25 May, 2024.

strong assumption that two name entities are available for training. However, it is not realistic that name entities are already provided.

To address the above challenges, we propose a prompt-based model with semantic knowledge augmentation (ZS-SKA) to perform zero-shot RTE. We first implement data augmentation by a word-level sentence translation to generate augmented instances with unseen relations from training instances with seen relations. We follow a new generation rule introduced in Sec. 3.2.1 to generate high-quality augmented instances for training in zero-shot settings. Note that ZS-SKA is trained only on labeled data from seen classes and augmented data generated from seen classes.

Secondly, inspired by prompt-tuning on pre-trained language models (Schick and Schütze, 2021a,b), we design the prompts based on the knowledge graph to integrate semantic knowledge to generally infer the features of unseen relations using patterns learned from seen relations. For the prompt design, we consider semantic knowledge information, including relation descriptions, super-class of relations and name entities, and a general knowledge graph to effectively learn the unseen relations. Instead of using the real label word directly in the prompt template, we automatically search a set of appropriate label words based on the knowledge graph for each label. The weight of each appropriate label word is calculated based on its semantic knowledge information in Sec. 3.2.2. We calculate the distance between each appropriate label with the true label itself to help denoise the set of appropriate label words. Then, we construct virtual label words in the prompt by weighted averaging all appropriate label word candidates.

Finally, we apply prototypical networks (Snell et al., 2017) to compute a prototype representing each relation. Each prototype is the mean vector of embedded and augmented sentences with prompts belonging to one relation. Euclidean distance is calculated between query sentence embeddings with prototypes to predict relations. A distance threshold explored in the validation set is applied to determine the number of relation triplets. For name entity predictions, we use a name entity extractor to recognize different types of entities. Then, sorted entity types are compared with the super-class of name entities in the prompt to determine the final relation triplets. The contributions:

- We propose a prompt-based model with semantic knowledge augmentation (ZS-SKA) to extract triplets with unseen relations under the zero-shot setting. Unlike some previous works, ZS-SKA considers semantic information from different granularities and does not rely on other large models with additional training.
- We present a new word-level sentence translation rule to generate augmented instances with unseen relations from instances with seen relations. The augmented sentences are then used as the training sets for unseen relations.
- We propose prompts for training based on an external knowledge graph to integrate semantic knowledge information learned from seen relations. We construct weighted virtual label words for the mask in the prompt template instead of using the actual label sets.
- We demonstrate that ZS-SKA significantly outperforms state-of-the-art methods for relation extraction with unseen relations under the ZSL setting on three public datasets.

2. Related Work

2.1. Prompt Learning in NLP

With the development of Generative Pre-trained Transformer 3 (GPT-3) (Brown et al., 2020), prompt-based learning has received considerable attention. Language prompts have been proven to be effective in downstream tasks leveraging pre-trained language models (Trinh and Le, 2018; Davison et al., 2019; Petroni et al., 2019). Human-designed prompts have achieved promising results in few-shot learning for sentiment classification (Schick and Schütze, 2021a,b). To avoid labor-intensive prompt design, studies explore prompts that are generated automatically (Shin et al., 2020; Jiang et al., 2020; Gao et al., 2021). However, most of the studies focus on supervised or few-shot learning on text classification (Fei et al., 2022; Hu et al., 2021; Han et al., 2021b; Gu et al., 2021), event detection (Li et al., 2022b), relation classification (Han et al., 2021b; Chen et al., 2021b) and name entity recognition (Li et al., 2022a; Huang et al., 2022; Ma et al., 2021). Inspired by these works, we explore prompt-based zero-shot learning in RTE.

2.2. Zero-shot Relation Classification

Relation classification is the problem of classifying relations given two name entities within a sentence. Most existing works rely on sufficient human-labeled data or noisy labeled data by distant supervision. When no training instances are available, some studies use zero-shot relation classification to extract unseen relations. This is typically done using question-answering models by listing questions that define the relation's slot values (Levy et al., 2017; Cetoli, 2020). Some studies formulate relation extraction as a text entailment task (Obamuyide and Vlachos, 2018). Some

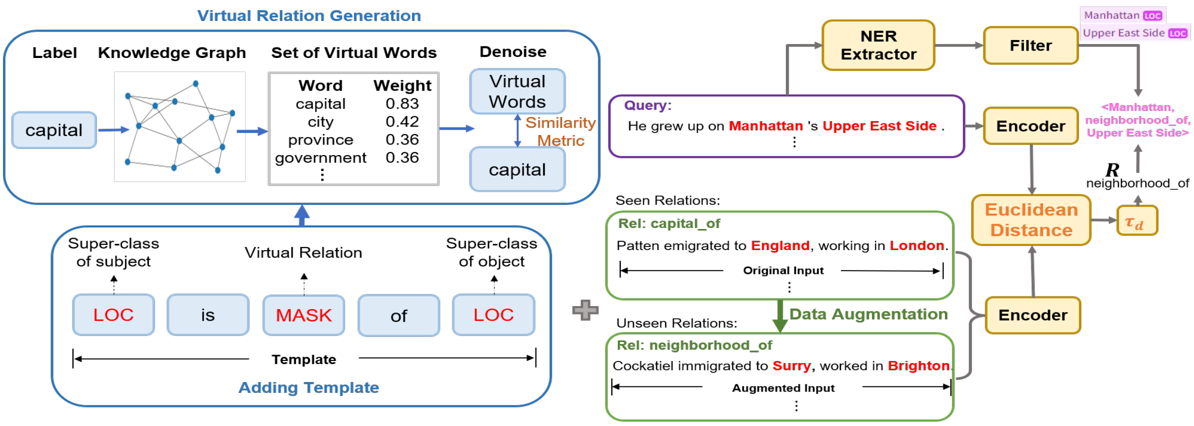

Figure 2: ZS-SKA overall architecture with components explained in Sec. 3.2.

studies utilize the accessibility of the relation descriptions to get the information for unseen relations (Obamuyide and Vlachos, 2018; Qin et al., 2020; Gong and Eldardiry, 2021; Chen and Li, 2021; Sainz et al., 2021; Liu et al., 2022a). However, these models only utilize class names semantic information, losing the connections between relations. Other studies focus on establishing the connection between relations with knowledge graph (Li et al., 2020) or contrastive learning (Wang et al., 2022). Nevertheless, all these works only focus on relation classification instead of relation triplet extraction. In the real world, it is not practical and realistic that two name entities are provided for training. Therefore, we focus on a more complex and realistic task: extracting both name entities and their relations.

2.3. Zero-shot Relation Triplet Extraction

Current approaches focusing on zero-shot relation triplet extraction (RTE) require large computing resources and additional deep-learning models. For example, RelationPrompt requires to fine-tune a pre-trained GPT-2 (124M parameters) as the relation generator (Chia et al., 2022). PCRED needs to train the entity boundary detection module by four neural network layers to get the possible boundaries for name entities in the triplet (Lan et al., 2022). ZETT views relation extraction as a template-filling problem, fine-tuning T5 to get the ranking score for potential triplets (Kim et al., 2022). However, ZETT can not discriminate against similar relations because the connections of different relations are lost. Inspired by data augmentation from knowledge graph in text classification (Zhang et al., 2019; Chen et al., 2021a) and prompt-based few-shot learning (Hu et al., 2021), we propose a prompt-based zero-shot RTE framework (ZS-SKA) incorporating external knowledge from the knowledge graph. Different from these existing works, the data augmentation module and name entity recognition module in ZS-SKA do not require any additional training. ZS-SKA can better catch the connections between relations due to the incorporated knowledge graph in the prompt template.

3. Methodology

In this section, we introduce the overall framework as shown in Figure 2 of ZS-SKA.

3.1. Problem Definition

To do zero-shot relation extraction, we adopt the problem setting in (Chia et al., 2022; Gong and Eldardiry, 2021) for zero-shot relation triplet extraction and setting in (Chen and Li, 2021) for zero-shot relation classification. We also conduct ablation experiments following the zero-shot definition in (Wenpeng Yin and Roth, 2019) which is a generalized zero-shot setting where partial labels are unseen. Given labeled instances belonging to a set of seen classes S, a model $M : X \to Y$ is learned, where $Y = Se \cup U$; U is the unseen class. For zero-shot RTE, let $R_s = \{r_s^1, \cdots, r_s^m\}$ and $R_u = \{r_u^1, \cdots, r_u^n\}$ denote the sets of seen and unseen relations, where $m = |R_s|$ and $n = |R_u|$ are the number of relations in the two disjoint sets, i.e., $R_s \cap R_u = \emptyset$. The set of relations R is predefined. The input of the training set consists of (1) seen relations R_s with input sentences X_i, (2) unseen relations R_u with super-class $S(e_{i_h})$, $S(e_{i_t})$ super-class of two name entities, and (3) external knowledge graph G. Here super-class is the hypernym of the item. For example, location (LOC) is the super-class of New York. The output of the model is a triplet in the format of <head (e_{i_h}), relation (r), tail (e_{i_t})> or a set of triplets if X_i contains multiple triplets. Our goal is to train a zero-shot relation triplet extraction model M to (1) learn the representations of both seen and unseen relations,

(2) predict new triplet $<e_{i_h}, r_u, e_{i_t}>$, where the relation r_u is not seen during the training phase. M is learned by minimizing the semantic distance between the embedding of the input and relation representations built from the knowledge graph G. For the zero-shot RC, the only difference with zero-shot RTE is that the two name entities e_{i_h} and e_{i_t} are known information with the input sentence X_i. Therefore, the model only needs to predict r_u given X_i with e_{i_h} and e_{i_t}.

3.2. Semantic Knowledge Augmentation

3.2.1. Data Augmentation

To enable the model to detect unseen relations without labeled training instances, we first do data augmentation by translating a sentence from its original seen relation to a new unseen relation using an analogy. In the word level, we adopt 3CosMul (Levy and Goldberg, 2014), where we use the top 10 similar words to return, to get the candidates of new words w_u:

$$w_u = \underset{x \in V}{argmax} \frac{cos(x, r_u) \cdot cos(x, w_s)}{cos(x, r_s) + \epsilon} \tag{1}$$

where V is the vocabulary set, $cos(\cdot)$ is the cosine similarity, r_u is the unseen relation, r_s is the seen relation, w_s is the word in seen class and ϵ is a small number to prevent division by zero.

Algorithm 1: Sentence Generation for Unseen Relations

Input : sentence $x_i = [w_1^i, \cdots, w_n^i]$, two name entities e_{i_h} and e_{i_t}, original relation label sets R_s, target unseen relation label r_u
Output : sentence x_i^u with relation r_u

for $r_s \in R_s$ **do**
  if $S(r_u) == S(r_s)$ *and* $S(e_u) == S(e_s)$ **then**
    for $w \in x_i$ **do**
      if *is_valid_pos(w)* **then**
        $w_u = 3CosMul(w, r_u, r_s)$
        x_i^u.append(w_u)
      else
        x_i^u.append(w)
  else
    Continue
return x_i^u

At the sentence level, we follow Algorithm 1 to translate a sentence of relation r_s into a new sentence of relation r_u. To be more specific, we translate all nouns, verbs, adjectives, and adverbs in the seen sentence to a new sentence. We do the translation when the super-class of r_s and the super-class of two corresponding name entities in r_s are the same as the super-class of r_u and the super-class of two related name entities in r_u. If the number of r_s that conforms to the rules is larger than one, we take all the translated sentences and randomly select the same number as other seen relations to make a balanced training set.

3.2.2. Prompts from Knowledge Graph

For relation extraction, the core issue is to extract the possible triplets from all aspects and granularities. For zero-shot tasks, we design prompts used as training instances to help train the model because there is no real training data available. We construct prompts based on the external knowledge graph ConceptNet (Speer and Havasi, 2013), a knowledge graph that connects words and phrases of natural language with labeled edges, for zero-shot relation extraction. Nodes in ConceptNet are entities, and edges connecting two nodes are semantic relations between the entities. Because of the relation extraction task, we wrap the input sequence with a *template*, which is a piece of natural language text. To be more specific, we build prompts as '$S(e_{i_h})$ is [MASK] of $S(e_{i_t})$'. We consider different locations of prompts such as before and after the input sentence. There is a similar performance, so we put the prompts after each input sentence. The [MASK] here is a virtual label word r_v representing the relation between $S(e_{i_h})$ and $S(e_{i_t})$. Unlike using real words, we build the virtual label word that can primarily represent the relation in each sentence. Instead of building a virtual label word by simply using the mean vector of the top_k high-frequency words (Ma et al., 2021), we build our virtual label word based on a knowledge graph using the following strategy.

We firstly represent a relation r as five sets of nodes in ConceptNet by processing the class label r_c, class hierarchy $S(r_c)$, class description $D(r_c)$ and hierarchy of two name entities $S(e_{i_h})$ and $S(e_{i_t})$. We consider whether a word w_i is related to the members of the five sets above within K hops or not. The value of K is determined through the grid search on the validation set. For each of the five sets above, we consider v_1 (whether w_i is a node in G in that set), v_2 (whether w_i's neighbor is a node in G), v_3 (number of neighbors of w_i in G). The above values associated with each set demonstrate the semantic distance of w_i and the corresponding set. The detailed construction of virtual label r_v is shown in Algorithm 2.

3.3. Model Architecture and Training

3.3.1. Instance Encoder

We first tokenize and lemmatize all words in a sentence. Two special tokens [CLS] and [SEP] are appended to the first and last positions, respectively. Then BERT (Devlin et al., 2019) is used to

Algorithm 2: Virtual Label Generation

Input : word w_i, relation r_c, threshold τ_s, number of hop K, Knowledge Graph G, number of virtual label n
Output : virtual label r_v
for $w_i \in V$ **do**
  if $\frac{w_i \cdot r_c}{|w_i| \times |r_c|} \geq \tau_s$ **then**
    v_1 = 0, v_2, v_3, v_{ave} = []
    if $w_i \in G$ **then**
      v_1 = 1
    else
      v_1= 0
    for $k \in K$ **do**
      hops = find_neighbors(w_i) $\in G$
      if *hops* **then**
        v_2.append(any(hops))
        v_3.append(sum(hops))
        v_{ave}.append(mean(hops))
      else
        v_2,v_3,v_{ave}.append(0)
      $\alpha_{w_i} = \frac{\sum v}{Dim(v)}$
  else
    Continue
  $\gamma_v = \frac{\alpha_{w_i} \cdot E(w_i) + \cdots + \alpha_{w_n} \cdot E(w_n)}{\sum \alpha}$
return r_v

generate the contextual representation for each token w_i. Because relation is not only related to the original name entities in augmented sentences generated by data augmentation, we have not used any position embeddings to show the positions of e_{i_h} and e_{i_t}. Let h_i represent the hidden state of the input sentence. We use $CNN(\cdot)$ ReLU and a max-pooling layer $max(\cdot)$, to derive the representation:

$$h_i = max(ReLU(CNN(x_i))) \tag{2}$$

where x_i is the tokenized input sentence:

$$x_i = w_{i-\frac{n-1}{2}}, \cdots, w_{i+\frac{n-1}{2}} \tag{3}$$

We obtain the hidden state vectors of prompts h_p:

$$h_p = E(S(e_{i_h})) \oplus E(r_v) \oplus E(S(e_{i_t})) \tag{4}$$

where $E(\cdot)$ is the embedding function, $S(\cdot)$ is the super-class of the input word and r_v is the virtual label embedding. The final representation for each instance is the concatenation of h_i and h_p.

3.3.2. Name Entity Extractor

ZS-SKA includes a name entity recognition encoder to extract name entities given the input X_i with predicted relation r_i. A fine-tuned BERT [1] is implemented to recognize different types of entities. Each entity e_i will be assigned a possibility score $score_i$ with entity types $T(e_i)$ (i.e. B-PER, I-LOC).

$$< e_i, T(e_i), score_i >= E_{NER}(X_i) \tag{5}$$

[1] https://huggingface.co/dslim/bert-base-NER

where E_{NER} is the name entity extractor based on BERT. Based on the predicted relation sets R_i, super-classes of the possible name entities $S(e_{i_h})$ and $S(e_{i_t})$ can be accessed from the prompt templates. Then, we filter out the entities whose types are different from the super-classes in the prompt template for each relation:

$$E_i = \{e_i | T(e_i) = S(e_i), score_i >= \tau_e\} \tag{6}$$

where E_i is the possible entity sets after the filter, $S(e_i)$ is the super-class of the target relation in the prompt template and τ_e is the threshold for the possibility of name entity types.

3.3.3. Model Training

The objective of training ZS-SKA is to minimize the distance between each instance embedding $h_i \oplus h_p$ and the prototype c_i embedding representing each learned relation (different colors in prototypes representation in Figure 2). Instead of using a softmax layer to classify seen relations and unseen relations, we adopt prototypical networks to compute a prototype for each relation after the BERT-CNN encoder. Each prototype is the average instance embeddings belonging to one relation:

$$c_i = \frac{1}{N} \sum_{i=1}^{N} f_\phi(h_i \oplus h_p) \tag{7}$$

where c_i represents the prototype for each relation, f_ϕ is the BERT-CNN encoder, h_i is the representation for each original or augmented sentence and p_i is denotes the prompt embeddings introduced in Sec. 3.2.2. The probabilities of the relations in R_s and R_u for a query instance x is calculated as:

$$p_\phi(y = r_i | x) = \frac{exp(-d(f_\phi(h_i \oplus h_p), c_i))}{\sum_{j=1}^{|R|} exp(-d(f_\phi(h_i \oplus h_p), c_j))} \tag{8}$$

where $d(.)$ is the Euclidean distance for two vectors. For multiple zero-shot RTE, we set a distance threshold τ_d to determine the number of possible unseen relations. During the inference phase, ZS-SKA predicts the relation set R_i by comparing the normalized distance $d(x_i)$ with the threshold τ_d:

$$R_i = \{r_i | softmax(d(x_i)) < \tau_d, r_i \in R\} \tag{9}$$

where $d(.)$ is the Euclidean distance. The final distance threshold τ_d for the testing phase is chosen by the threshold value that has the best performance in the evaluation phase. For zero-shot RTE, the final relation triplets are the combination of E_i from Equ. 6 and R_i from Equ. 9. For zero-shot RC, only R_i is provided for the result.

Table 1: Results for Zero-Shot Relation Triplet Extraction.

#unseen relations	Model	FewRel Single Acc.	FewRel Multi Pre.	FewRel Multi Rec.	FewRel Multi F1	Wiki_ZSL Single Acc.	Wiki_ZSL Multi Pre.	Wiki_ZSL Multi Rec.	Wiki_ZSL Multi F1
m=5	TabSeq (Wang and Lu, 2020)	11.82	15.23	1.91	3.40	14.47	43.68	3.51	6.29
	NoGen (Chia et al., 2022)	11.49	9.45	**36.74**	14.57	9.05	15.58	**43.23**	22.26
	RelPrompt (Chia et al., 2022)	22.27	20.80	24.32	22.34	16.64	29.11	31.00	30.01
	$\text{ZETT}_{\text{T5-small}}$ (Kim et al., 2022)	26.34	31.12	30.01	30.53	20.24	31.62	32.41	31.74
	$\text{ZETT}_{\text{T5-base}}$ (Kim et al., 2022)	30.71	38.14	30.58	33.71	21.49	35.89	28.38	31.74
	PCRED (Lan et al., 2022)	22.67	43.91	34.97	**38.93**	18.40	38.14	36.84	37.48
	ZS-SKA (ours)	**32.86**	**57.50**	26.24	36.04	**44.00**	**66.70**	27.24	**38.68**
m=10	TabSeq (Wang and Lu, 2020)	12.54	28.93	3.60	6.37	9.61	45.31	3.57	6.4
	NoGen (Chia et al., 2022)	12.40	6.40	**41.70**	11.02	7.10	9.63	**45.01**	15.70
	RelPrompt (Chia et al., 2022)	23.18	21.59	28.68	24.61	16.48	30.20	32.31	31.19
	$\text{ZETT}_{\text{T5-small}}$ (Kim et al., 2022)	23.07	25.52	29.61	27.28	14.37	19.86	27.71	22.83
	$\text{ZETT}_{\text{T5-base}}$ (Kim et al., 2022)	27.79	30.65	32.44	31.28	17.16	24.49	26.99	24.87
	PCRED (Lan et al., 2022)	24.91	30.89	29.90	30.39	22.30	27.09	39.09	32.00
	ZS-SKA (ours)	**34.03**	**60.48**	23.22	**33.28**	**26.40**	**45.38**	29.27	**35.30**
m=15	TabSeq (Wang and Lu, 2020)	11.65	19.03	1.99	3.48	9.20	**44.43**	3.53	6.39
	NoGen (Chia et al., 2022)	10.93	4.61	**36.39**	8.15	6.61	7.25	**44.68**	12.34
	RelPrompt (Chia et al., 2022)	18.97	17.73	23.20	20.08	16.16	26.19	32.12	28.85
	$\text{ZETT}_{\text{T5-small}}$ (Kim et al., 2022)	21.08	16.20	23.22	18.90	10.74	14.96	19.31	16.79
	$\text{ZETT}_{\text{T5-base}}$ (Kim et al., 2022)	**26.17**	22.50	27.09	24.39	12.78	19.45	23.31	21.21
	PCRED (Lan et al., 2022)	25.14	27.00	23.55	25.16	**21.64**	25.37	33.80	28.98
	ZS-SKA (ours)	23.86	**37.29**	19.13	**25.29**	20.26	31.23	27.20	**29.19**

4. Experiments

We conduct several experiments with ablation studies on three public datasets: FewRel, Wiki-ZSL, and NYT to show that our proposed model outperforms other existing state-of-the-art models, and our proposed model is more robust compared with the other models in the zero-shot learning tasks.

4.1. Evaluation Settings

4.1.1. Dataset

In our experiments, we evaluate our model [2] over three widely used datasets: FewRel (Han et al., 2018), Wiki-ZSL (Chen and Li, 2021) and NYT (Riedel et al., 2010). FewRel and Wiki-ZSL are two balanced datasets and NYT is an unbalanced dataset. The statistics of FewRel, Wiki-ZSL, and NYT datasets are shown in Table 6. We provide a more detailed description in Sec. 7.1.

4.1.2. Baselines and Evaluation Metrics

For RTE in ZSL, we compare our proposed model ZS-SKA with six SOTA zero-shot RTE models: **TableSequence** (Wang and Lu, 2020), **NoGen** (Chia et al., 2022), **RelationPrompt** (Chia et al., 2022), **ZETT** (Kim et al., 2022) with two sizes of T5 models (T5-small and T5-base), and **PCRED** (Lan et al., 2022). For the zero-shot RC task, We compare our proposed model to eight existing RC models on all three public datasets to evaluate the model's ability to detect unseen relations. For clean FewRel and Wiki-ZSL datasets, we compare our model with **CNN** (Zeng et al., 2014), **Bi-LSTM** (Zhang et al., 2015), **Attentional Bi-LSTM** (Zhou et al., 2016), **R-BERT** (Wu and He, 2019), **ESIM** (Chen et al., 2017), **CIM** (Rocktäschel et al., 2016), **ZS-BERT** (Chen and Li, 2021), and **NoGen** (Chia et al., 2022). The eight baselines above are reported by (Chen and Li, 2021) and (Chia et al., 2022). We also compare the robustness of our model with the most SOTA re-implemented **ZS-BERT**, **NoGen** and **RelationPrompt**. For noisy NYT dataset, we compare our model with the re-implemented **CDNN** (Zeng et al., 2014), **REDN** (Li and Tian, 2020) and **ZSLRC** (Gong and Eldardiry, 2021). The evaluation metric for a single RTE is Accuracy (Acc.) because each sentence only includes one gold triplet. The evaluation metrics for multiple RTE are Precision (Pre.), Recall (Rec.), and F1-score (F1), because there are at least two gold triplets in the testing set. For RC evaluation, we also use Precision, Recall, and F1-score, similar to those used for the above baselines.

4.1.3. Parameter Settings

We follow the experiment settings as (Chia et al., 2022) and (Chen and Li, 2021) to enable zero-shot RTE and zero-shot RC tasks. We randomly select m unseen relations and remove all the instances related to these m relations in the training set to ensure that these m relations have not appeared

[2] Code: `https://github.com/gjiaying/ZS-SKA`

Table 2: RC results with different m values on NYT.

	m=15			m=30		
	Precision	Recall	F1	Precision	Recall	F1
CDNN	27.94 $\pm$ 0.52	44.10 $\pm$ 0.44	33.72 $\pm$ 0.46	10.17 $\pm$ 0.86	25.62 $\pm$ 0.71	14.23 $\pm$ 0.77
REDN	66.52 $\pm$ 0.47	65.47 $\pm$ 0.62	66.98 $\pm$ 0.55	57.19 $\pm$ 0.60	56.80 $\pm$ 0.70	56.99 $\pm$ 0.62
ZSLRC	96.06 $\pm$ 0.35	93.84 $\pm$ 0.21	93.59 $\pm$ 0.31	94.81 $\pm$ 0.38	**90.46$\pm$0.22**	89.76 $\pm$ 0.29
ZS-SKA	**96.23 $\pm$ 0.08**	**94.68 $\pm$ 0.12**	**94.42 $\pm$ 0.11**	**95.91 $\pm$ 0.20**	90.38 $\pm$ 0.32	**91.27 $\pm$ 0.28**

in training data. m is varied to examine how performance is affected. More details of hyperparameter settings are discussed in Sec. 7.2 and Table 7.

4.2. Results and Discussion

4.2.1. Zero-shot Relation Triplet Extraction

Main Results Table 1 shows the results of both single and multiple RTE on FewRel and Wiki_ZSL in ZSL. The results of our proposed model are reported by the average of five runs. For single RTE, we observe that ZS-SKA significantly outperforms other baselines when m=5 and m=10. From Table 1, ZS-SKA demonstrates better performance on Wiki_ZSL than on FewRel, as it shows a greater increase in accuracy compared to the strongest baseline. This indicates that our proposed model is more robust and effective on the dataset with more classes as Wiki_ZSL has 113 relations and FewRel has 80 relations in all. For multiple RTE, ZS-SKA consistently achieves the best F1 score in all settings, except when m=5 on the FewRel dataset. Compared to other baselines, ZS-SKA achieves a relatively high precision score, resulting in a better F1 score. Downstream applications for RTE, such as building knowledge graphs using the extracted triplets, require high-quality data. A high precision and relatively low recall performance may result in missing some gold labels (i.e. missing links for the knowledge graph). However, a high recall and low precision performance (i.e. NoGen achieves the best recall performance in all settings) means that the model returns many results, but most of its predicted labels are incorrect compared to the gold labels. This may introduce much noise (i.e. a noisy dataset) to downstream tasks. We conjecture that the high precision performance of ZS-SKA is due to setting a relatively smaller distance threshold in Sec. 3.3.3 for determining the number of relation triplets. In future work, a dynamic threshold could be added to adjust to different datasets.

4.2.2. Zero-shot Relation Classification

ZS-SKA also supports the zero-shot relation classification task by providing two name entities in the training set. Recognizing unseen relations is mainly supported by semantic knowledge augmentation in Sec. 3.2. Therefore, we carry out relation classification experiments on NYT, FewRel and Wiki_ZSL datasets to better evaluate zero-shot ability of ZS-SKA.

Table 3: RC results (m=15) on Wiki-ZSL/FewRel.

	Pre.	Rec.	F1
CNN	14.58/14.17	17.68/20.26	15.92/16.67
BiLSTM	16.25/16.83	18.94/27.62	17.49/20.92
$\text{BiLSTM}_{\text{att}}$	16.93/16.48	18.54/26.36	17.70/20.28
R-BERT	17.31/16.95	18.82/19.37	18.03/18.08
ESIM	27.31/29.15	29.62/31.59	28.42/30.32
CIM	29.17/31.83	30.58/33.06	29.86/32.43
ZS-BERT	34.12/35.54	34.38/38.19	34.25/36.82
NoGen	**54.45/66.49**	29.43/40.05	37.56/**49.38**
ZS-SKA	41.78/45.03	**40.50/51.86**	**39.30**/46.99

Results on Unbalanced Dataset The experiment results on unbalanced dataset NYT by varying m unseen relations are shown in Table 2. To make fair comparisons, we use the same splitted NYT dataset and follow the same threshold schema provided by (Gong and Eldardiry, 2021). We remove the data augmentation module and only implement the prompts generated through the knowledge graph as similar side information in ZSLRC model. Apparently, the proposed ZS-SKA achieves a substantial gain in precision, recall, and F1-score over other baselines on the NYT dataset. When the number of unseen relations in the testing set becomes larger, the superiority of ZS-SKA gets more significant and robust. Such results indicate the effectiveness of leveraging prompts using virtual labels constructed from the knowledge graph instead of using keywords learned from the distribution of training data on the noisy dataset in (Gong and Eldardiry, 2021).

Results on Balanced Datasets The evaluation results of zero-shot RC on Wiki-ZSL and FewRel are shown in Table 3. The results of all baselines are reported by (Chen and Li, 2021; Chia et al., 2022) and the result of ZS-SKA is reported by the average of five different random seeds. For fair comparison, we compare our proposed model with baselines that do not require any training process for additional models such as the generator (GPT-2). Obviously, ZS-SKA significantly outperforms other existing models on both balanced datasets for recall value. Besides, ZS-SKA has the best

Table 4: Ablation study (F1) over ZS-SKA on Wiki-ZSL with different percentages of unseen relations.

	10%	20%	30%	40%	50%
ZS-BERT	58.31	19.59	17.63	11.79	9.52
NoGen	42.72	26.20	18.20	12.28	8.69
RelPrompt	**67.91**	**50.02**	**36.51**	22.13	12.94
Ours_{Aug}	41.44	33.57	26.00	22.04	16.00
$\text{Ours}_{\text{Prompts}}$	46.59	36.20	27.76	19.32	13.72
$\text{Ours}_{\text{Top2 freq}}$	41.10	33.68	28.49	22.40	18.60
$\text{Ours}_{\text{Top5 freq}}$	40.90	34.85	28.84	22.68	18.25
$\text{Ours}_{\text{ActLabel}}$	42.06	34.89	28.85	22.93	18.43
$\text{Ours}_{\text{OnlyBert}}$	37.35	31.80	25.80	20.75	16.51
Ours_{All}	40.93	35.99	28.97	**24.64**	**19.27**

F1 performance on Wiki-ZSL. The performance improvement indicates that semantic knowledge augmentation is competitively more beneficial for recognizing unseen relations than only incorporating text description of relations.

4.3. Analysis

4.3.1. Ablation Study

To evaluate the robustness and effectiveness of the zero-shot ability of ZS-SKA, we conduct an ablation study on Wiki-ZSL by removing different modules from ZS-SKA.The zero-shot setting is followed by the definition that partial relations are unseen in the testing set (Wenpeng Yin and Roth, 2019). This setting is more competitive because all classes (including both seen and unseen relations) exist in the testing set. Different with the experiments of specific m values, this is a 113 class classification experiment, including different percentages of unseen relations, which is more related to the real-world scenario. From Table 4, we observe that ZS-SKA is more robust when increasing the proportions of unseen relations. The performance drops drastically for ZS-BERT and NoGen when more unseen relations appear. RelationPrompt is more stable than ZS-BERT and NoGen. But the performace also drops a lot starting from 40% of unseen relations. Though instances generated by data augmentation for unseen relations may include noise, the models with data augmentation can be more robust when large percentages of unseen relations exist in the testing set. We also implement models with different ways to construct the prompt such as using top k frequency words, actual label itself to evaluate the virtual label construction in ZS-SKA. Virtual label construction is more effective when 20% or more of unseen relations exist. It is because prompts constructed by virtual labels contain the semantic information of unseen relations, which shortens the distance between the query sentence of an unseen relation with the unseen relation prototype.

4.3.2. Case Study

Virtual Label Construction Figure 3 shows an example of ranking the top ten components of the constructed virtual label before denoising and after denoising. The virtual labels shown in Figure 3 are generated by Algorithm 2. The red words are irrelevant to the relation 'religion_of'. After we refine the virtual label sets using the distance metric, these irrelevant words are filtered out in our virtual label sets, removing the noise in the knowledge graph.

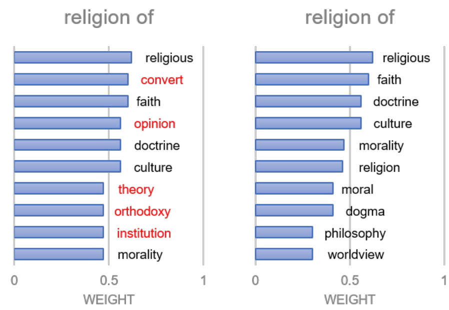

Figure 3: Denoising in virtual label construction.

Name Entity Extractor Figure 4 shows an example of how ZS-SKA uses the name entity extractor with super-classes information to extract relation triplets. ZS-SKA includes two steps for RTE. First, unseen relations are predicted by semantic knowledge augmentation. Based on the predicted relations, super-class information such as 'LOC' and 'PER' can be accessed from the prompt template. Second, the NER extractor is implemented to extract the types of name entities. For example, the relation 'birthplace' happens between 'PER' and 'LOC' according to the template. The filter in the NER extractor selects 'Boyd' in 'PER' and 'Boston' in 'LOC'. Similarly, relation 'capital_of' only happens between 'LOC' and 'LOC', so the filter in the NER extractor selects 'Boston' and 'Massachusetts' in 'LOC'. Note that all locations in 'LOC' in Figure 4 are ranked based on the possibility score. Then, the predicted relations and entities extracted by the NER extractor construct the final relation triplets.

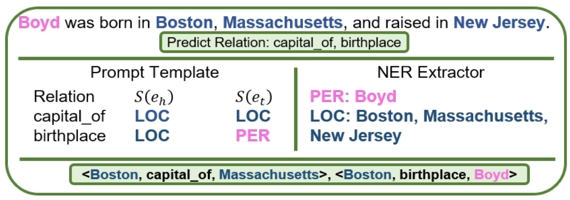

Figure 4: Example of using Name Entity Extractor to extract relation triplets.

Table 5: Examples of sentence generation from seen relations by data augmentation. Words in red are name entities for each sentence. $S(\cdot)$ denotes the super-class of the relation or name entities.

Relation r	$S(r)$	$S(e_1)$	$S(e_2)$	Sentence
place_of_birth	location	person	location	Jessica (born in Manchester) is a British track and field athlete who competes in the heptathlon.
place_of_death	location	person	location	Johnson (died in Liverpool) is a Military track and field athlete who competed in the decathlon.
residence	location	person	location	Mansion (resided in Villa) is a Colonial residence and peri alumnus who dominates in the decathlon.
country	location	location	location	Rich (retired in Arsenal) is a European track and field athlete who competes in the decathlon.
educated_at	action	person	organization	Jess (motivate in Liverpool) is a British aims and professional athlete who educated in the decathlon.

Data Augmentation Table 5 shows an example of the augmented data following the translating rule on the Wiki-ZSL dataset. The relation 'place_of_birth' is a seen class, and the other four relations are from unseen classes. We use data augmentation to generate augmented training instances for these unseen relations. We observe that if the super-class of both the relation and two name entities are the same, the generated sentences have a good quality with the name entities having unseen relations. If the super-class of the relation or two name entities of unseen relation is different from that of the seen relation, though the generated sentences contain the tone of the unseen relation (words in blue), the original two name entities do not have the target unseen relation. For example, the generated sentence of relation 'country' can be explained that Arsenal is from a European country, but such relation is lost between the two name entities 'Rich' and 'Arsenal'. Therefore, we follow the rule of using the relation and name entities from the same super-class with that of unseen relations to generate high-quality augmented instances for training in ZSL.

4.3.3. Hyperparameter Sensitivity

We examine how some hyperparameters, including threshold τ_s for denoising virtual label sets and the number of virtual labels n, affect ZS-SKA's performance. By fixing m = 15 and varying τ_s and n, the results are exhibited in Figure 5. We find that τ_s and n affect the noisy dataset more than the balanced dataset. We think that because both τ_s and n are used for removing noise and getting more related semantic information in prompts, the noise in prompts may impact more on noisy datasets because they are more sensitive to the noise.

It is obvious that τ_s does have an impact on the performance. If the threshold τ_s is between 0.5 and 0.6, it achieves the best performance on all three public datasets. This is reasonable that when τ_s is too low, most connected nodes in the knowledge graph are used to construct virtual label words, and more noise may be obtained from the relation. In contrast, when τ_s gets too high, some highly related nodes are filtered out to construct virtual labels. We also find that increasing the number of related words n to construct virtual labels can achieve better performance. It is reasonable because, including more nodes (words) from the knowledge graph to construct the virtual label representing the relation, more semantic information is contained, leading to a shorter distance.

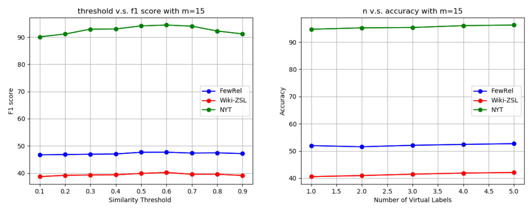

Figure 5: Effects on varying threshold τ and number of virtual labels n on three datasets.

5. Conclusion and Future Work

We propose a ZS-SKA utilizing semantic knowledge augmentation to extract unseen relation triplets with no labeled data available for training to tackle with zero-shot RTE. The experiments show that with augmented instances, prompts generated through a knowledge graph, and a NER extractor with prompts, ZS-SKA outperforms other SOTA zero-shot RTE models. We have also conducted extensive experiments to study different aspects of ZS-SKA, from ablation study, and case study to hyperparameter sensitivity, and demonstrate the effectiveness and robustness of our proposed model. In future work, we plan to explore: (1) Different ways of instance generation and prompt designs for semantic augmented data. (2) Better approaches for constructing virtual labels in the prompt template. (3) More SOTA data augmentation techniques to generate data for zero-shot tasks to further improve the performance.

6. Bibliographical References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Alberto Cetoli. 2020. Exploring the zero-shot limit of FewRel. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1447–1451, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Chih-Yao Chen and Cheng-Te Li. 2021. Zs-bert: Towards zero-shot relation extraction with attribute representation learning. In *Proceedings of 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-2021)*.

Qi Chen, Wei Wang, Kaizhu Huang, and Frans Coenen. 2021a. Zero-shot text classification via knowledge graph embedding for social media data. *IEEE Internet of Things Journal*, pages 1–1.

Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017. Enhanced lstm for natural language inference. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, Vancouver. ACL.

Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2021b. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. *CoRR*, abs/2104.07650.

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 2022*, pages 45–57, Dublin, Ireland. Association for Computational Linguistics.

Joe Davison, Joshua Feldman, and Alexander Rush. 2019. Commonsense knowledge mining from pretrained models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1173–1178, Hong Kong, China. Association for Computational Linguistics.

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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Bowen Dong, Yuan Yao, Ruobing Xie, Tianyu Gao, Xu Han, Zhiyuan Liu, Fen Lin, Leyu Lin, and Maosong Sun. 2020. Meta-information guided meta-learning for few-shot relation classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1594–1605, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Chunliu Dou, Shaojuan Wu, Xiaowang Zhang, Zhiyong Feng, and Kewen Wang. 2022. Function-words adaptively enhanced attention networks for few-shot inverse relation classification. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 2937–2943. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Yu Fei, Zhao Meng, Ping Nie, Roger Wattenhofer, and Mrinmaya Sachan. 2022. Beyond prompting: Making pre-trained language models better zero-shot learners by clustering representations. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8560–8579, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Tianyu Gao, Adam Fisch, and Danqi Chen. 2021. Making pre-trained language models better few-shot learners. In *Association for Computational Linguistics (ACL)*.

Tianyu Gao, Xu Han, Zhiyuan Liu, and Maosong Sun. 2019. Hybrid attention-based prototypical networks for noisy few-shot relation classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6407–6414.

Jiaying Gong and Hoda Eldardiry. 2021. *Zero-Shot Relation Classification from Side Information*, page 576–585. Association for Computing Machinery, New York, NY, USA.

Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang. 2021. Ppt: Pre-trained prompt tuning for few-shot learning.

Jiale Han, Bo Cheng, and Wei Lu. 2021a. Exploring task difficulty for few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2605–2616, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2021b. Ptr: Prompt tuning with rules for text classification.

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*, pages 4803–4809, Brussels, Belgium. Association for Computational Linguistics.

Shengding Hu, Ning Ding, Huadong Wang, Zhiyuan Liu, Juanzi Li, and Maosong Sun. 2021. Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification.

Yucheng Huang, Kai He, Yige Wang, Xianli Zhang, Tieliang Gong, Rui Mao, and Chen Li. 2022. COPNER: Contrastive learning with prompt guiding for few-shot named entity recognition. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2515–2527, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Pere-Lluís Huguet Cabot and Roberto Navigli. 2021. REBEL: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2370–2381, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. 2020. How can we know what language models know. *Trans. Assoc. Comput. Linguistics*, 8:423–438.

Bosung Kim, Hayate Iso, Nikita Bhutani, Estevam Hruschka, and Ndapa Nakashole. 2022. Zero-shot triplet extraction by template infilling. *arXiv preprint arXiv:2212.10708*.

Yuquan Lan, Dongxu Li, Hui Zhao, and Gang Zhao. 2022. Pcred: Zero-shot relation triplet extraction with potential candidate relation selection and entity boundary detection. *arXiv preprint arXiv:2211.14477*.

Omer Levy and Yoav Goldberg. 2014. Linguistic regularities in sparse and explicit word representations. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pages 171–180, Ann Arbor, Michigan. Association for Computational Linguistics.

Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 333–342, Vancouver, Canada. Association for Computational Linguistics.

Cheng Li and Ye Tian. 2020. Downstream model design of pre-trained language model for relation extraction task. *CoRR*, abs/2004.03786.

Dongfang Li, Baotian Hu, and Qingcai Chen. 2022a. Prompt-based text entailment for low-resource named entity recognition. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1896–1903, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Haochen Li, Tong Mo, Hongcheng Fan, Jingkun Wang, Jiaxi Wang, Fuhao Zhang, and Weiping Li. 2022b. KiPT: Knowledge-injected prompt tuning for event detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1943–1952, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Juan Li, Ruoxu Wang, Ningyu Zhang, Wen Zhang, Fan Yang, and Huajun Chen. 2020. Logic-guided semantic representation learning for zero-shot relation classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2967–2978, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Fangchao Liu, Hongyu Lin, Xianpei Han, Boxi Cao, and Le Sun. 2022a. Pre-training to match for unified low-shot relation extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5785–5795, Dublin, Ireland. Association for Computational Linguistics.

Yang Liu, Jinpeng Hu, Xiang Wan, and Tsung-Hui Chang. 2022b. A simple yet effective relation information guided approach for few-shot relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 757–763, Dublin, Ireland. Association for Computational Linguistics.

Ruotian Ma, Xin Zhou, Tao Gui, Yiding Tan, Qi Zhang, and Xuanjing Huang. 2021. Template-free prompt tuning for few-shot ner.

Youmi Ma, Tatsuya Hiraoka, and Naoaki Okazaki. 2022. Joint entity and relation extraction based on table labeling using convolutional neural networks. In *Proceedings of the Sixth Workshop on Structured Prediction for NLP*, pages 11–21, Dublin, Ireland. Association for Computational Linguistics.

Abiola Obamuyide and Andreas Vlachos. 2018. Zero-shot relation classification as textual entailment. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 72–78, Brussels, Belgium. Association for Computational Linguistics.

Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. 2019. Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473, Hong Kong, China. Association for Computational Linguistics.

Pengda Qin, Xin Wang, Wenhu Chen, Chunyun Zhang, Weiran Xu, and William Yang Wang. 2020. Generative adversarial zero-shot relational learning for knowledge graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8673–8680.

Feiliang Ren, Longhui Zhang, Shujuan Yin, Xiaofeng Zhao, Shilei Liu, Bochao Li, and Yaduo Liu. 2021. A novel global feature-oriented relational triple extraction model based on table filling. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2646–2656, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Haopeng Ren, Yi Cai, Xiaofeng Chen, Guohua Wang, and Qing Li. 2020. A two-phase prototypical network model for incremental few-shot relation classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1618–1629, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases*, pages 148–163, Berlin, Heidelberg. Springer Berlin Heidelberg.

Tim Rocktäschel, Edward Grefenstette, Karl Moritz Hermann, Tomás Kociský, and Phil Blunsom. 2016. Reasoning about entailment with neural attention. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label verbalization and entailment for effective zero and few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1199–1212, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Timo Schick and Hinrich Schütze. 2021a. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 255–269. Association for Computational Linguistics.

Timo Schick and Hinrich Schütze. 2021b. It's not just size that matters: Small language models are also few-shot learners. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2339–2352. Association for Computational Linguistics.

Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. 2020. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 4222–4235. Association for Computational Linguistics.

Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Daniil Sorokin and Iryna Gurevych. 2017. Context-aware representations for knowledge base relation extraction. In *Proceedings of the 2017 Con-*

ference on Empirical Methods in Natural Language Processing*, pages 1784–1789, Copenhagen, Denmark. Association for Computational Linguistics.

Robyn Speer and Catherine Havasi. 2013. *ConceptNet 5: A Large Semantic Network for Relational Knowledge*, pages 161–176. Springer Berlin Heidelberg, Berlin, Heidelberg.

Trieu H. Trinh and Quoc V. Le. 2018. A simple method for commonsense reasoning. *CoRR*, abs/1806.02847.

Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1706–1721, Online. Association for Computational Linguistics.

Shusen Wang, Bosen Zhang, Yajing Xu, Yanan Wu, and Bo Xiao. 2022. RCL: Relation contrastive learning for zero-shot relation extraction. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2456–2468, Seattle, United States. Association for Computational Linguistics.

Jamaal Hay Wenpeng Yin and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In *EMNLP*.

Shanchan Wu and Yifan He. 2019. Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, page 2361–2364, New York, NY, USA. Association for Computing Machinery.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. 2019. Integrating semantic knowledge to tackle zero-shot text classification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long Papers)*, Minneapolis, USA. Association for Computational Linguistics.

Senhui Zhang, Tao Ji, Wendi Ji, and Xiaoling Wang. 2022. Zero-shot event detection based on ordered contrastive learning and prompt-based prediction. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2572–2580, Seattle, United States. Association for Computational Linguistics.

Shu Zhang, Dequan Zheng, Xinchen Hu, and Ming Yang. 2015. Bidirectional long short-term memory networks for relation classification. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pages 73–78, Shanghai, China.

Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for entity and relation extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 50–61, Online. Association for Computational Linguistics.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 207–212, Berlin, Germany. Association for Computational Linguistics.

Hao Zhu, Yankai Lin, Zhiyuan Liu, Jie Fu, Tat-Seng Chua, and Maosong Sun. 2019. Graph neural networks with generated parameters for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1331–1339, Florence, Italy. Association for Computational Linguistics.

7. Appendix

7.1. Dataset Description

Table 6: The statistics of each dataset.

	#instances	#relations	avg. Len.
FewRel	56,000	80	24.95
Wiki-ZSL	94,383	113	24.85
NYT	134,152	53	38.81

- ***FewRel* (Han et al., 2018).** The FewRel dataset is a human-annotated balanced few-shot RC dataset consisting of 80 types of relations, each of which has 700 instances.

- ***Wiki-ZSL (Chen and Li, 2021).*** The Wiki-ZSL dataset is a subset of Wiki-KB (Sorokin and Gurevych, 2017), which filters out both the 'none' relation and relations that appear fewer than 300 times.

- ***NYT (Riedel et al., 2010).*** The NYT dataset was generated by aligning Freebase relations with the New York Times. There are 53 possible relations in total. It is an unbalanced noisy dataset.

7.2. Parameter Settings

Table 7: Parameter Settings

Parameter	Value
Word Embedding Dimension	768
Hidden Layer Dimension	300
Sentence Max Length	128
Convolutional Window Size	3
Batch Size	4
Initial Learning Rate α	0.01
Weight Decay	10^{-5}
Number of Hops K	1
Similarity Threshold τ_s	0.6
Distance Threshold τ_d	0.05
NER Threshold τ_e	0.5
Number of Virtual Label n	5

For the hyperparameter and configuration of ZS-SKA, we implement ZS-SKA with PyTorch and optimize it with an SGD optimizer. The initial learning rate is selected via the grid search within the range of $\{1e-1, 1e-2, 1e-3, 1e-4\}$ for minimizing the loss, the cosine similarity threshold is selected from 0 to 1 with step size 0.1. The distance threshold for determining the number of triplets in a given sentence is set to 0.05, which is explored in the validation set. NER threshold is selected from 0.1 to 0.9 with a step size of 0.2. Table 7 shows other parameters. We follow the early stopping strategy when selecting the model for testing. The model is evaluated on the validation set every 50 epochs. The time for training is around 6 hours depending on the computing resources. GPU with 16G memory is required for training.

7.3. Limitations

Given the progress made to date with the work we propose in this paper, we view the following current limitations as some opportunities to build on in future work. First, data augmentation is based on word-level transformation. With the development of generation models, more state-of-the-art data augmentation techniques can be implemented to generate data for zero-shot tasks to further improve the performance. Second, the proposed prompt method depends on information from a fixed knowledge graph, which means it can not deal with the scenario if the unseen label is an out-of-vocabulary word. We have not considered this scenario because all classes from the three public datasets are well-known words or phrases. In future work, to get prompt information when the class word does not exist in the knowledge graph, we will consider directly using label descriptions or text generation models such as GPT-2 to generate label explanations.

7.4. Ethical Considerations

Data Bias Our proposed model ZS-SKA is specifically intended for zero-shot relation triplet extraction or zero-shot relation classification tasks. We perform experiments on three public datasets. However, the model's performance may be subject to bias when applied to other datasets with significantly different distributions or in new domains. Therefore, we advise exercising caution when assessing the generalizability and fairness of the model.

Computing Cost Our model needs the use of GPU training, which imposes a computational burden. We acknowledge that this burden has an adverse environmental impact on carbon emissions. Specifically, our research requires 6 hours of training on a single GPU card for each task. In total, we have 6 $\times$ 5 (each task has 5 runs) $\times$ 15 tasks (6 tasks in RTE, 4 tasks in RC, and 5 tasks in ablation study) = 450 hours of training, resulting in 112.5lbs of carbon dioxide.