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

Anonymous ACL submission

Abstract

In relation classification, recognizing unseen (new) relations for which there are no training instances is a challenging task. 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 sentence translation rule and generate augmented instances with unseen relations from instances with seen relations using that new rule. We design prompts 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 show that ZS-SKA outperforms state-of-the-art methods under the zero-shot scenarios. Our experimental results also demonstrate the effectiveness and robustness of ZS-SKA.

1 Introduction

Relation classification aims to infer the semantic relation between a pair of entities in a sentence. However, existing approaches based on supervised learning (Zhu et al., 2019; Li and Tian, 2020) or few-shot learning (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 definition of a more generalized zero-shot setting that partial classes are new (Wenpeng Yin and Roth, 2019).

Zero-shot relation classification aims to classify relations of name entities in a sentence that are absent from the learning stage. Existing approaches on zero-shot relation classification still have limitations. First, some models perform zero-shot relation classification by listing questions that define the relation's slot values (Levy et al., 2017). These models have a strong assumption that an excellent question-answering model is learned, and all values extracted from this model are correct. This is impractical in the real-world setting. 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 state-of-the-art models leverage side (auxiliary) information to tackle zero-shot tasks. They focus on class names/descriptions semantic information, losing the connection or relationships between seen relations and unseen relations (Gong and Eldardiry, 2021; Chen and Li, 2021).

To address the above challenges, we propose a prompt-based model with semantic knowledge augmentation (ZS-SKA) to perform zero-shot learning for relation classification. We first implement data augmentation by a word-level sentence translation to generate augmented instances with unseen relations from training instances with seen relations. The super-class of the triplet (subject, relation, object) for augmented instances is the same as the triplet of training instances. 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 a knowledge graph to integrate semantic knowledge to generally infer the features of unseen relations using patterns learned from seen relations. For 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 (augmented)sentences with prompts belonging to one relation. Euclidean distance is calculated between query sentence embeddings with prototypes to classify relations. The contributions of this paper can be summarized as follows:

- We propose a prompt-based model with semantic knowledge augmentation (ZS-SKA) to predict 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 complex models.
- 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 instances for unseen relations.
- We propose prompts based on an external knowledge graph to integrate semantic knowledge information learned from seen relations. We construct weighted virtual label words for mask in prompt template instead of using actual label sets.
- We demonstrate that ZS-SKA significantly outperforms state-of-the-art methods for predicting unseen relations under the ZSL setting on three public datasets. Results show the effectiveness and robustness of ZS-SKA.

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 proved 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 (Hu et al., 2021; Han et al., 2021; Gu et al., 2021), relation classification (Han et al., 2021; Chen et al., 2021b) and name entity recognition (Ma et al., 2021).

2.2 Zero-shot Relation Classification

Relation extraction is the problem of extracting semantic relations between two name entities within a given sentence. When no training instances are available, some studies use zero-shot relation classification to extract unseen relation types. This is typically done using question-answering models. In particular, by listing questions that define the relation's slot values (Levy et al., 2017; Cetoli, 2020). To avoid relying on question-answering models, some studies formulate relation extraction as a text entailment task and utilize the accessibility of the relation descriptions (Obamuyide and Vlachos, 2018; Qin et al., 2020; Gong and Eldardiry, 2021; Chen and Li, 2021). 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). Nevertheless, they miss the semantic information from name entities. Inspired by data augmentation from knowledge graph in text classification tasks (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 relation classification framework incorporating external knowledge from the knowledge graph.

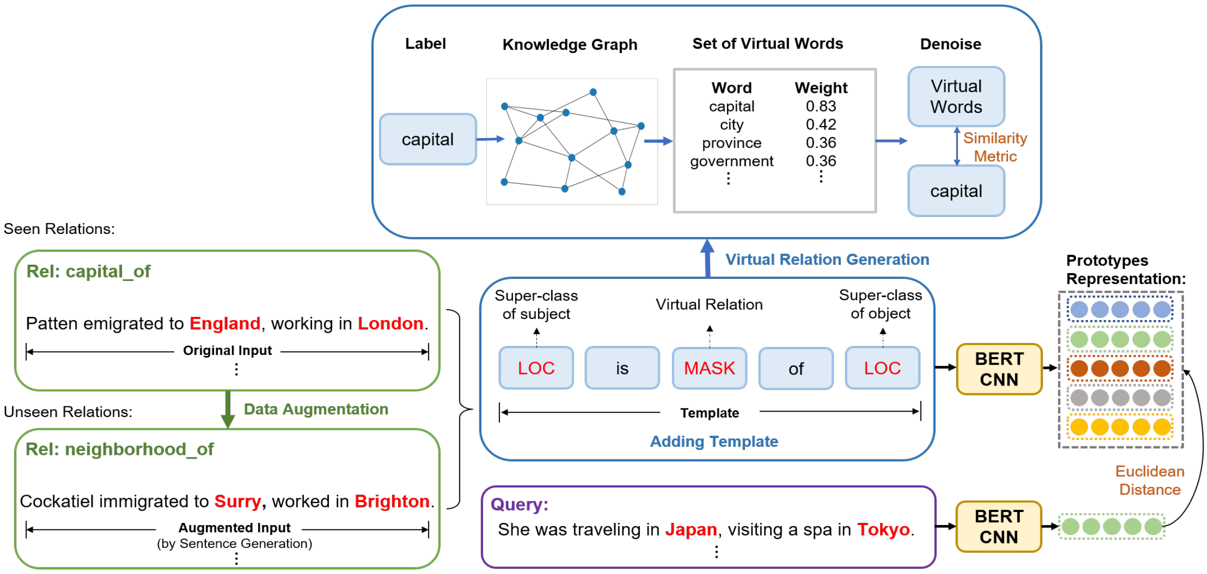


Figure 1: ZS-SKA architecture with components explained in Sec. 3.2.

3 Methodology

In this section, we introduce the overall framework as shown in Figure 1 of our proposed prompt-based ZS-SKA, starting with problem formulation.

3.1 Problem Definition

We follow the definition of zero-shot from (Wenpeng Yin and Roth, 2019) and the same settings from (Chen and Li, 2021; Gong and Eldardiry, 2021) to conduct our experiments. This is a more generalized zero-shot setting that partial labels are unseen. Given labeled instances belonging to a set of seen classes S, a model $M : X \rightarrow Y$ is learned, where $Y = S \cup U$; U is the unseen class.

For relation classification task, 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., $R_s \cap R_u = \emptyset$. Given the training set consisting of seen relations R_s with their corresponding input sentences X_i and name entities e_{i1} and e_{i2}, unseen relations R_u, super-class of name entities $S(e_{i1})$, $S(e_{i2})$, relates to R_u and external knowledge graph G. Our goal is to train a zero-shot relation classification model M to learn the representations of both seen and unseen relations. M is learned by minimizing the semantic distance between the embedding of the input and relation representations built from the knowledge graph.

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 adopts 3CosMul [1] (Levy and Goldberg, 2014) to get the candidates of new words w_u:

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

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

In 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 with 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 above 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.

[1]We use top 10 similar words to return.

Algorithm 1: Sentence Generation for Unseen Relations

Input : sentence $x_i = [w_1^i, \cdots, w_n^i]$, two name entities e_{i1} and e_{i2}, 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

Algorithm 2: Virtual Label Generation

Input : word w_i, relation r_c, threshold τ, number of hop K, Knowledge Graph G, number of virtual label n

Output : virtual label r_v

for $w_i \in V$ **do**
  if $\frac{w_i \cdot r_c}{|w_i| \times |r_c|} \geq \tau$ **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

3.2.2 Prompts from Knowledge Graph

For relation classification, the core issue is to extract the relations related to the two given name entities 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. From this perspective, we construct prompts based on 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 classification. Nodes in ConceptNet are entities, and edges connecting two nodes are semantic relations between the entities.

Because of the relation classification 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 '$S(e_{i1})$ is [MASK] of $S(e_{i2})$' [2]. The [MASK] here is a virtual label word r_v representing the relation between e_{i1} and e_{i2}. 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 $S(e_{i1})$ and $S(e_{i2})$. 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 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. Detailed construction of virtual label r_v is shown in Algorithm 2.

3.3 Model Architecture and Training

3.3.1 Instance Encoder

Figure 2 shows the architecture of the encoder used in this paper. We first tokenize and lemmatize all words in a sentence. Tow special tokens [CLS] and

[2] 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.

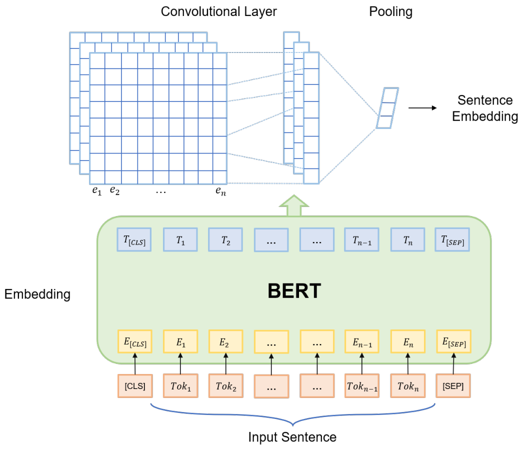


Figure 2: BERT-CNN Instance Encoder.

[SEP] are appended to the first and last positions, respectively. Then BERT (Devlin et al., 2019) is used to generate the contextual representation for each token w_i. Because the relation is not only related to the original two name entities in augmented sentences generated by data augmentation in Sec. 3.2.1, we have not used any position embeddings to indicate the positions of e_{i1} and e_{i2}. Let h_i represent the hidden state of the input sentence. We use a convolutional layer $CNN(\cdot)$ and a ReLU activation function, together with a max-pooling layer $max(\cdot)$, to derive the representation vector:

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

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 generated in Sec. 3.2.2 as:

$$h_p = E(S(e_{i1})) \oplus E(r_v) \oplus E(S(e_{i2})) \tag{4}$$

where $E(\cdot)$ denotes the embedding function, $S(\cdot)$ represents the super-class of the input word and r_v is the virtual label embedding introduced in Sec. 3.2.2. The final representation for each instance is the concatenation of h_i and h_p.

3.3.2 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 relation. 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 BERT-CNN encoder.

Table 1: 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

Each prototype is the average instance embeddings belonging to one relation:

$$c_i = \frac{1}{N}\sum_{i=1}^{N} f_\phi(h_i \oplus h_p) \tag{5}$$

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:

$$p_\phi(y = r_i|x) = \frac{exp(-d(f_\phi(h_i \oplus h_p), c_i))}{\sum_{j=1}^{|R|} exp(-d(f_\phi(h_i \oplus h_p), c_j))} \tag{6}$$

where $d(.)$ is Euclidean distance for two vectors.

4 Experiments

We conduct several experiments with ablation studies on three public datasets: FewRel (Han et al., 2018), Wiki-ZSL (Sorokin and Gurevych, 2017; Chen and Li, 2021) and NYT (Riedel et al., 2010) 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 zero-shot learning tasks.

4.1 Evaluation Settings

4.1.1 Dataset

In our experiments, we evaluate our model 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 1. We provide more detailed description in the Appendix.

4.1.2 Zero-shot Settings

We follow the experiment settings as (Chen and Li, 2021) to enable zero-shot relation classification 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 in training data. For hyperparameter and configuration of ZS-SKA, we implement ZS-SKA with PyTorch and optimize it with SGD optimizer. The initial learning rate is selected via 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. Table 5 in Appendix shows other parameters used in the experiment.

4.1.3 Baselines and Evaluation Metrics

We compare our proposed model to several state-of-the-arts models in zero-shot learning tasks. 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) and **ZS-BERT** (Chen and Li, 2021). The seven baselines above are reported by (Chen and Li, 2021). We also compare the robustness of our model with the most state-of-the-art re-implemented **ZS-BERT**. 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 metrics adopted in this paper are the Precision, Recall, and F1-score, similar to those used for the above baselines.

4.2 Results and Discussion

4.2.1 Main Results

Results on Balanced Datasets The evaluation results of zero-shot learning on Wiki-ZSL and FewRel are shown in Table 2. We compare our proposed model ZS-SKA with models reported by (Chen and Li, 2021). Obviously, ZS-SKA significantly outperforms other state-of-the-art models on both balanced datasets. Our proposed ZS-SKA outperforms a recently proposed method (ZS-BERT) by 6.9% precision, 5.7% recall, and 3.9% F1-score on Wiki-ZSL, 9.8% precision, 13.5% recall, and 10.2% F1-score on FewRel. The performance improvement indicates that semantic knowledge augmentation is competitively more beneficial for relation classification than only incorporating text description of relations. To compare the robustness of ZS-SKA with the strongest baseline ZS-BERT, we conduct further experiments with different percentages (varying m) of unseen relations in Sec. 4.2.2.

Results on Unbalanced Dataset The experiment results on unbalanced dataset NYT by varying m unseen relations are shown in Table 3. 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 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.

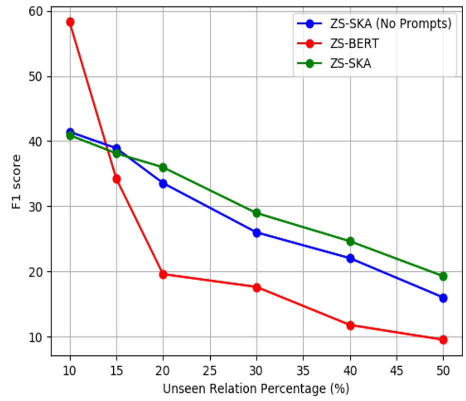


Figure 3: F1 scores of models with different proportions of unseen relations.

4.2.2 Ablation Study

To evaluate the robustness and effectiveness of modules in ZS-SKA, we conduct an ablation study on Wiki-ZSL by removing prompts from ZS-SKA to make comparisons with the most state-of-the-art model ZS-BERT and our model ZS-SKA. From Figure 3, we observe that our proposed model ZS-SKA is more robust when increasing the proportions of unseen relations in the testing set. ZS-BERT performs much better when only 10% of unseen relations exist. However, the performance drops drastically when more unseen relations appear. Removing the prompts in ZS-SKA performs slightly better when only 10% of unseen relations exist. Nevertheless, the performance drops more

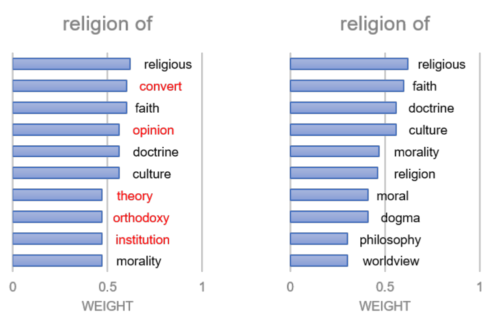
Table 2: Results with $m = 15$ on Wiki-ZSL and FewRel.

	Wiki-ZSL			FewRel		
	Precision	Recall	F1	Precision	Recall	F1
CNN (Zeng et al., 2014)	14.58	17.68	15.92	14.17	20.26	16.67
Bi-LSTM (Zhang et al., 2015)	16.25	18.94	17.49	16.83	27.62	20.92
Att Bi-LSTM (Zhou et al., 2016)	16.93	18.54	17.70	16.48	26.36	20.28
R-BERT (Wu and He, 2019)	17.31	18.82	18.03	16.95	19.37	18.08
ESIM (Chen et al., 2017)	27.31	29.62	28.42	29.15	31.59	30.32
CIM (Rocktäschel et al., 2016)	29.17	30.58	29.86	31.83	33.06	32.43
ZS-BERT (Chen and Li, 2021)	34.12	34.38	34.25	35.54	38.19	36.82
ZS-SKA	**41.03**	**40.12**	**38.13**	**45.34**	**51.67**	**47.02**

Table 3: Results with different m values on NYT.

	m=15			m=30		
	Precision	Recall	F1	Precision	Recall	F1
CDNN (Zeng et al., 2014)	27.94	44.10	33.72	10.17	25.62	14.23
REDN (Li and Tian, 2020)	66.52	65.47	66.98	57.19	56.80	56.99
ZSLRC (Gong and Eldardiry, 2021)	96.06	93.84	93.59	94.81	**90.46**	89.76
ZS-SKA	**96.23**	**94.68**	**94.42**	**95.91**	90.38	**91.27**

significantly than ZS-SKA. It is probably because prompts constructed by virtual labels contain the semantic information of unseen relations, which shorten the distance between the query sentence of an unseen relation with the prototype of such unseen relation.



(a) (b)

Figure 4: Examples of denoising in virtual label construction on FewRel and Wiki-ZSL datasets.

4.2.3 Case Study

Data Augmentation Table 4 shows an example of the augmented data following the the translating rule in Sec. 3.2.1 on Wiki-ZSL dataset. Relation ‘place_of_birth’ is a seen class, and the four relations ‘place_of_death’, ‘residence’, ‘country’ and ‘educated_at’ are from unseen classes. We follow the data augmentation method introduced in Sec. 3.2.1 to generate augmented training instances for these unseen relations. We observe that if the super-class of the relation and the super-class of the two name entities are the same, i.e. ‘place_of_death’ with ‘place_of_birth’, ‘residence’ with ‘place_of_birth’, the generated sentences have a good quality with the name entities having the unseen relations. If the super-class of the relation or super-class of the two name entities of unseen relation is different from that of the seen relation, i.e. ‘country’ with ‘place_of_birth’, ‘educated_at’ with ‘place_of_birth’, though the generated sentences contain the tone of the target (unseen) relation, such as the words in blue, the original two name entities do not have the target unseen relation. For example, the generated sentence of relation ’country’ in Table 4 can be explained that Arsenal is from a European country, but such relation is lost between the original 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 zero-shot settings.

Virtual Label Construction Figure 4 shows an example of the ranking top ten components of the constructed virtual label before denoising and after denoising. The virtual labels shown in Figure 4 are generated by Algorithm 2 proposed in Sec. 3.2.2.

Table 4: 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	act	person	org.	Jess (motivate in Liverpool) is a British aims and professional athlete who educated in the decathlon.

The red words in Figure 4 (a) 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.

4.2.4 Hyperparameter Sensitivity

We examine how some primary hyperparameters, including threshold τ for denoising virtual label sets and the number of virtual labels n in Algorithm 2 affect the performance of ZS-SKA. By fixing $m = 15$ and varying τ and n, the results in terms of F1 scores and Accuracy on NYT, FewRel and Wiki-ZSL datasets are exhibited in Figure 5. We find that parameters τ and n affect the noisy dataset more than the clean and balanced dataset. We conjecture that because both τ and n are used for removing noise and getting more related semantic information in prompts construction, the noise in prompts may impact more on noisy datasets because noisy datasets are more sensitive to the noise.

If the threshold τ is between 0.5 and 0.6, it achieves the best performance on all three public datasets. This is reasonable that when τ is too low, most connected nodes in the knowledge graph are used to construct virtual label words. Thus, when building the prompts for each relation, it is more likely to bring the noise to the relation class. In contrast, when τ gets too high, some highly related nodes are filtered out to construct virtual labels. We would suggest setting τ between 0.5 to 0.6 to derive satisfying results across datasets. As for the number of words n to construct virtual labels, we 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 information, more semantic knowledge information is contained, leading to a shorter distance between the query sentence embedding with the prototype constructed from the prompts.

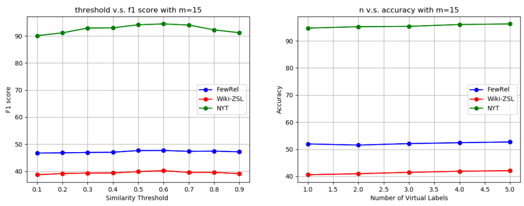


Figure 5: Effects on varying threshold τ and number of virtual labels n on NYT, FewRel and Wiki-ZSL datasets.

5 Conclusion and Future Work

In this paper, we propose a prompt-based ZS-SKA utilizing semantic knowledge augmentation to detect unseen relations with no corresponding labeled data available for training to tackle with zero-shot relation classification task. The experiments show that with augmented instances and prompts generated through a knowledge graph, ZS-SKA outperforms other state-of-the-art models under zero-shot learning. We have also conducted extensive experiments to study different aspects of ZS-SKA, from ablation study to hyperparameter sensitivity, and demonstrate the effectiveness and robustness of our proposed model. We plan to explore the following directions in future work: (1) Different ways of instance generation and prompt designs for semantic augmented data. (2) Better approaches for constructing virtual labels in the prompt template.

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.

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.

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.

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2021. 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.

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.

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.

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.

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

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.

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*.

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 Conference 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.

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.

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.

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.

A Dataset Description

In the following, we describe each dataset in detail.

- ***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 corpos (NYT). There are 53 possible relations in total. It is an unbalanced noisy dataset because all the relations have a different number of sentences.

B Parameter Settings

Table 5: 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
Number of Hops K	1
Similarity Threshold τ	0.6
Number of Virtual Label n	5