Zero-shot Relation Classification from Side Information

Jiaying Gong
Virginia Tech
Blacksburg, U.S.
gjiaying@vt.edu

Hoda Eldardiry
Virginia Tech
Blacksburg, U.S.
hdardiry@vt.edu

CIKM2021

Code:https://github.com/gjiaying/zslrc

2022.3.29 • ChongQing

Reported by Yang Peng

Artificial

1.Introduction

2.Method

3.Experiments

4. Conclusion

Introduction

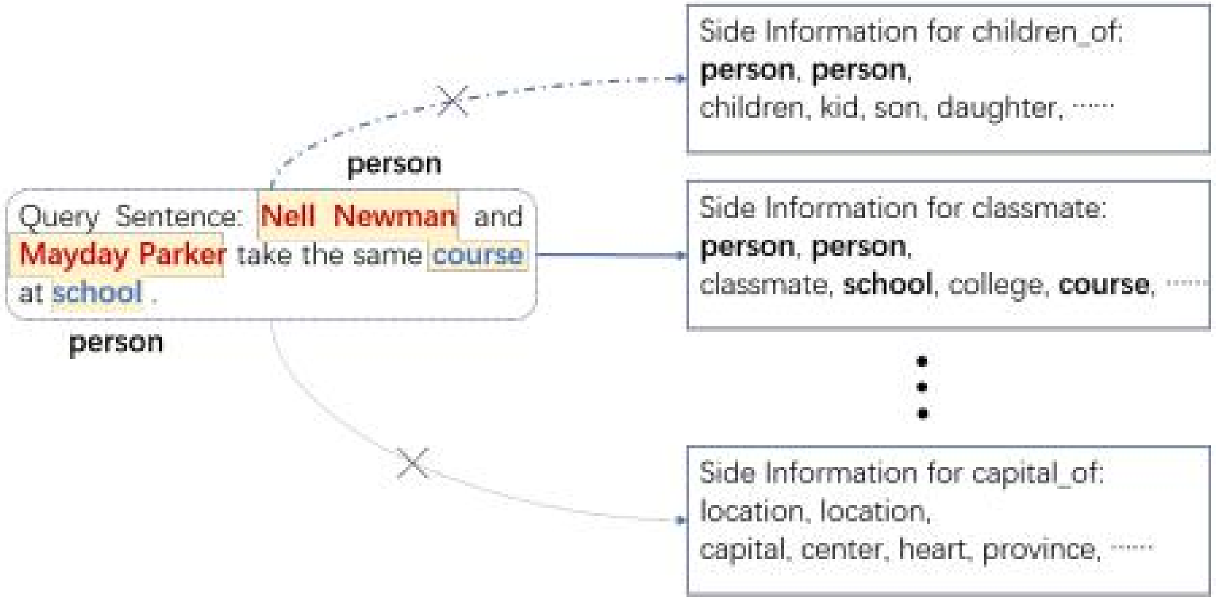


Figure 1: Example of relation classification based on side information.

Introduction

- We propose the first approach (ZSLRC) to enable zero-shot learning on relation classification without relying on other complex models that need to be learned and assumed to be 100% accurate.

- ZSLRC uses side information including labels, keywords, and hypernyms of name entities, and it has been shown that our model can perform competitively using the weighted side information.

- We modify prototypical networks to recognize new relations in addition to recognized previously known relations.

Method

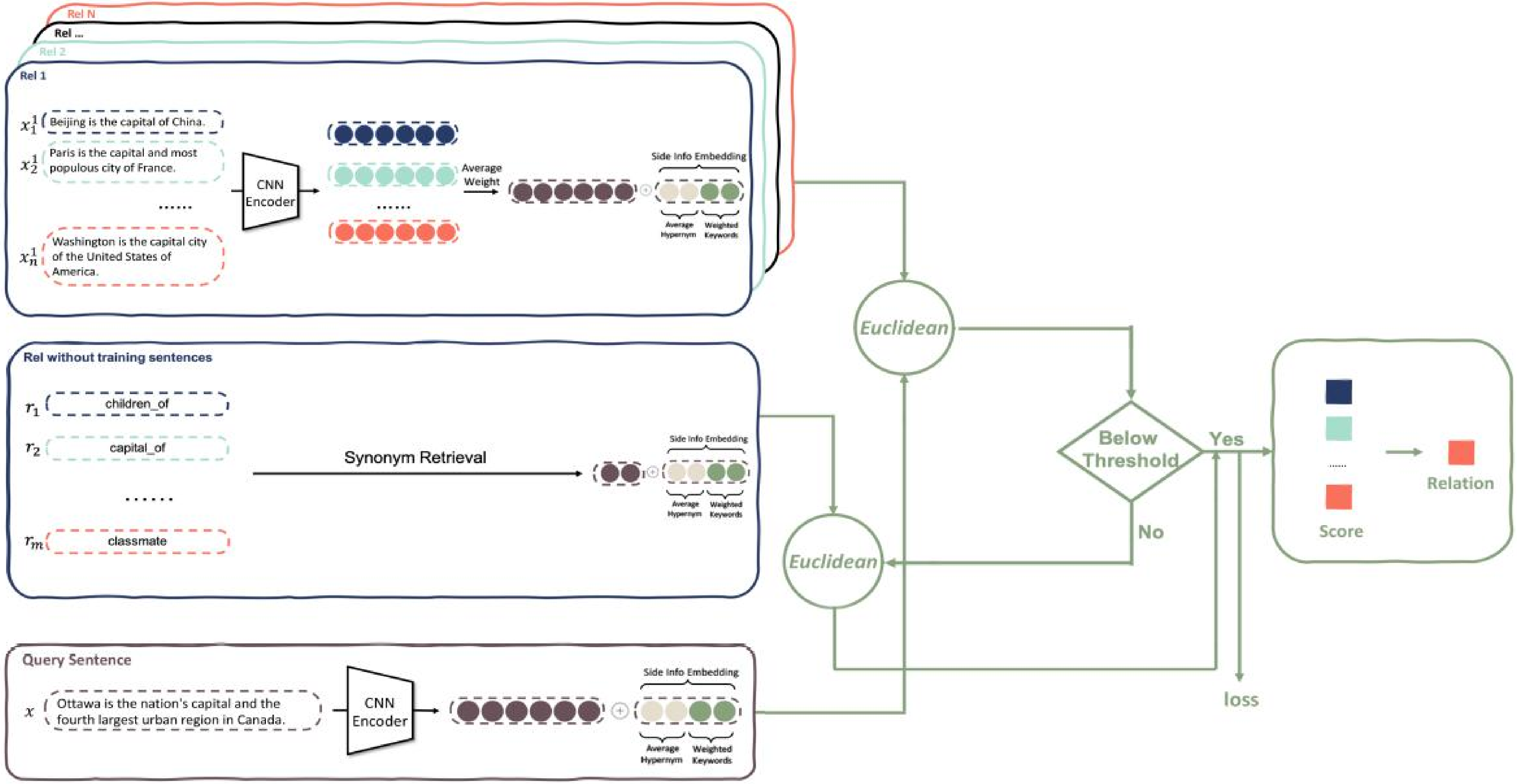
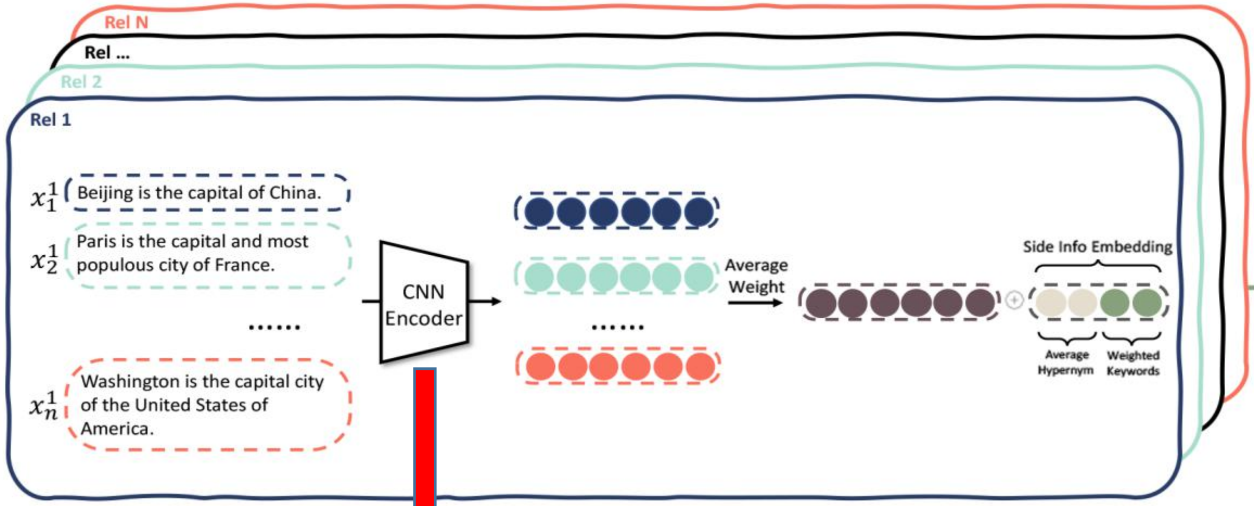


Figure 2: Model of Zero-shot Learning for Relation Classification (ZSLRC)

Method



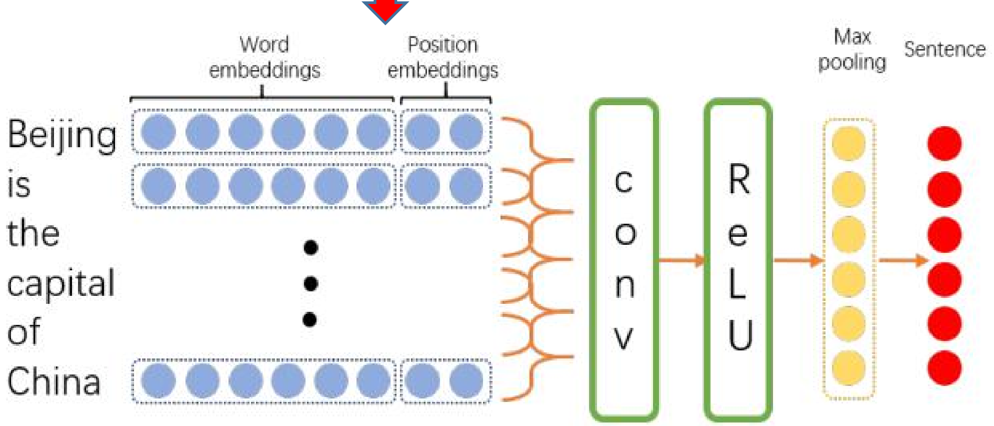


Figure 3: CNN Encoder

inputs sentences $\{x_1, x_2, x_3, \cdots x_n\}$

For each word $w \in S = \{w_1, w_2, \cdots w_n\}$ $S \in \mathbb{R}^{s \times d}$,

$$\hat{w}_i = w_i \oplus p_{i1} \oplus p_{i2} \tag{1}$$

s is the sentence length and $d = d_w + d_p \times 2$

Instance embedding as follows:

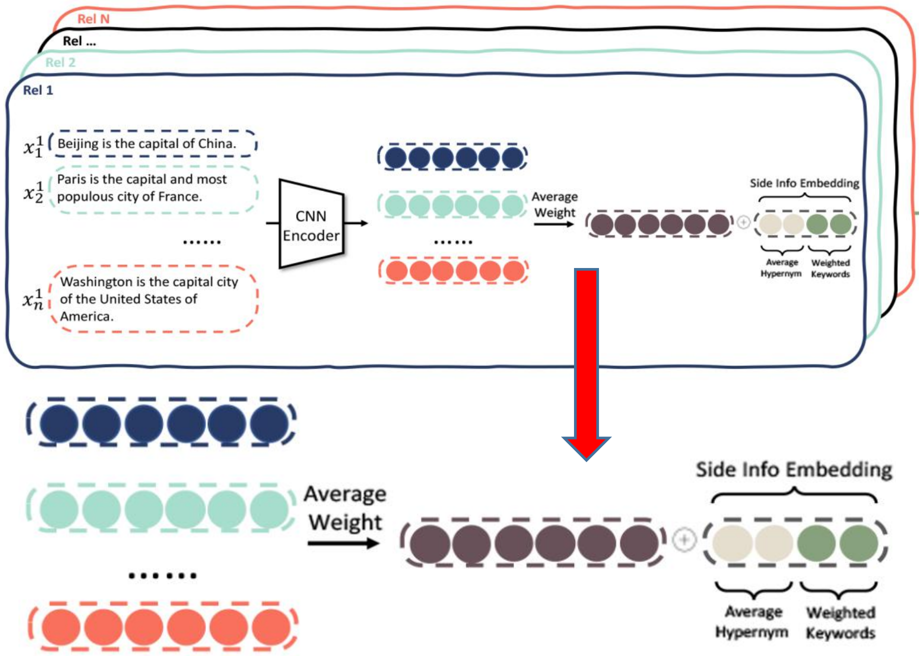
$$x_i = CNN(w_{i-\frac{n-1}{2}}, \cdots, w_{i+\frac{n-1}{2}}) \tag{2}$$

$$\hat{x_i} = max(0, x_i) \tag{3}$$

$$[s]_j = max\{[\hat{x_1}]_j, \cdots, [\hat{x_n}]_j\} \tag{4}$$

$[\cdot]_j$ is the j-th value of a vector.

Method



final prototype including side information for each relation can be expressed as follows:

$$c_i{}' = \begin{cases} r \oplus si_h \oplus si_r \oplus si_k & r \neq 0 \\ si_h \oplus si_r \oplus si_s & r = 0 \end{cases} \tag{5}$$

Each prototype is the mean vector of embedded sentences

$$c_i = \frac{1}{N}\sum_{i=1}^{N} f_\phi(x_i) \tag{6}$$

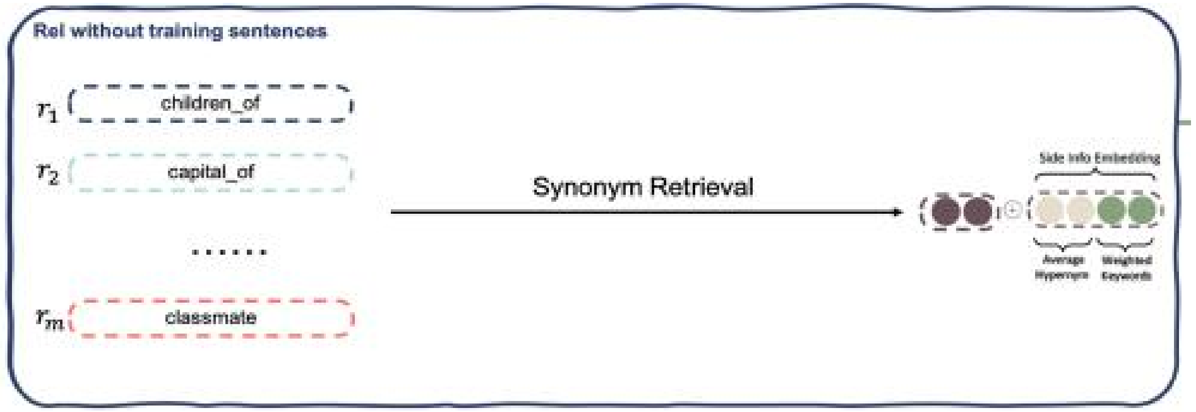
The equation of side information embedding si is as follows:

$$si = f(\frac{h_1 + h_2}{2}) \oplus f(k_1) \oplus \cdots \oplus f(k_n) \oplus K \tag{7}$$

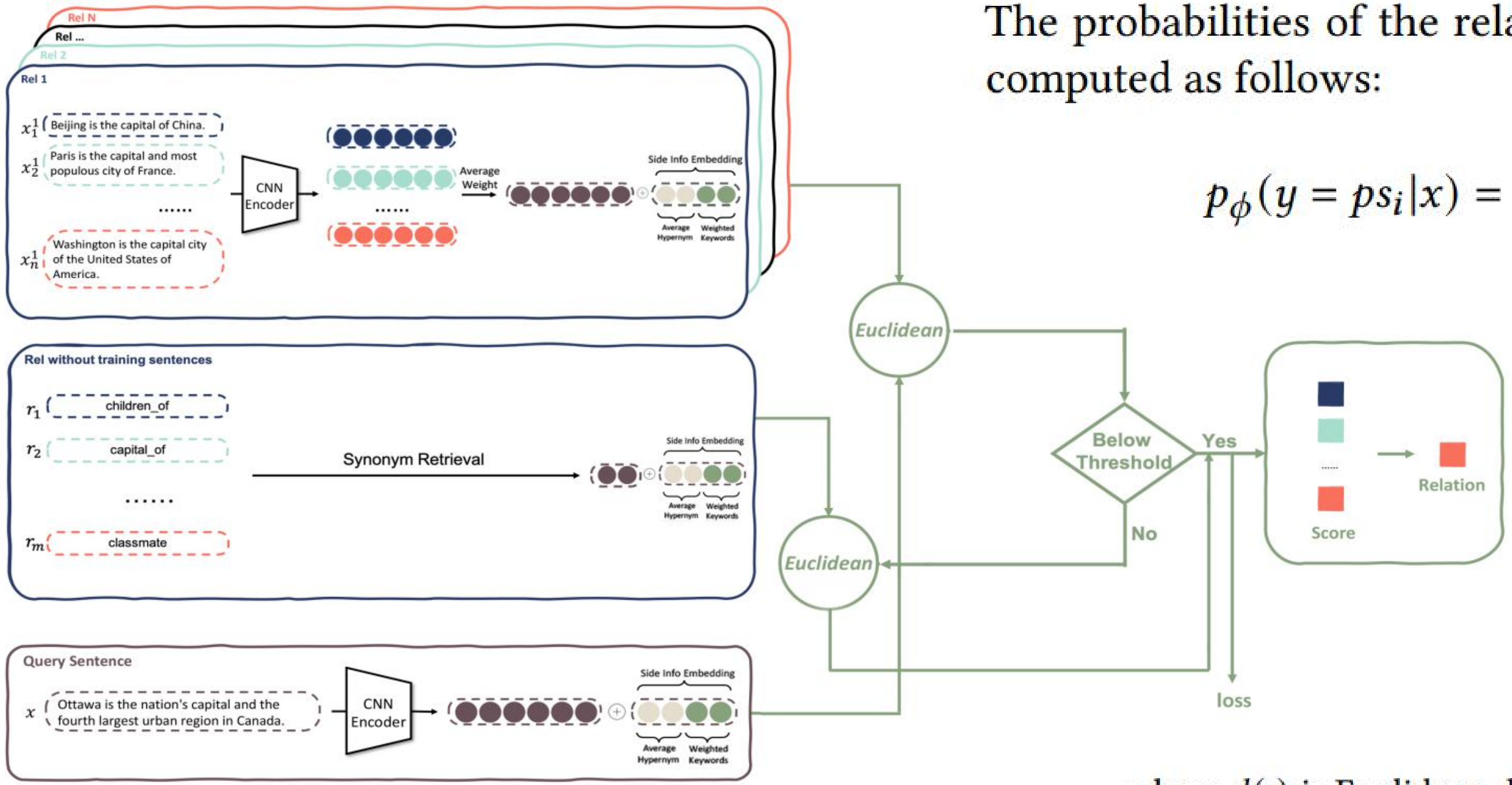
$$K = \sum_{m-n}^{m}(\frac{\alpha_i}{\sum_{i=m-n}^{m} \alpha_i} f(k_i)) \tag{8}$$

$$\alpha_i = \frac{count(k,s)}{size(s)} \cdot log(\frac{N}{sentence(k,S)}) \tag{9}$$

$$ps_i = c_i \oplus si_i \tag{10}$$



Method



The probabilities of the relations in $\Re$ for a query instance x is computed as follows:

$$p_\phi(y = ps_i|x) = \frac{exp(-d(f_\phi(x), ps_i))}{\sum_{ps'_i} exp(-d(f_\phi(x), ps'_i))} \tag{11}$$

where $d(.)$ is Euclidean distance function as below:

$$d(f_\phi(x), ps_i) = \sqrt{\sum_{i=1}^{n}(ps_i - f_\phi(x))^2} \tag{12}$$

Experiments

Table 1: Parameter Settings

Parameter	Value
Word Embedding Dimension d_w	50
Position Embedding Dimension d_p	5
Side Information Embedding Dimension d_{si}	300
Hidden Layer Dimension d_h	800
Convolutional Window Size n	3
Batch Size	1
Initial Learning Rate α	0.01
Weight Decay	10^{-5}
Threshold t	2e-08

Experiments

Table 2: Results of different models on NYT (%). Our re-implementation is marked by ∗.

Model	Precision	Recall	F1
CDNN* [43]	46.4	52.7	45.8
REDN [19]	95.1	94.0	94.6
ZSLRC	**98.1**	**97.9**	**97.6**

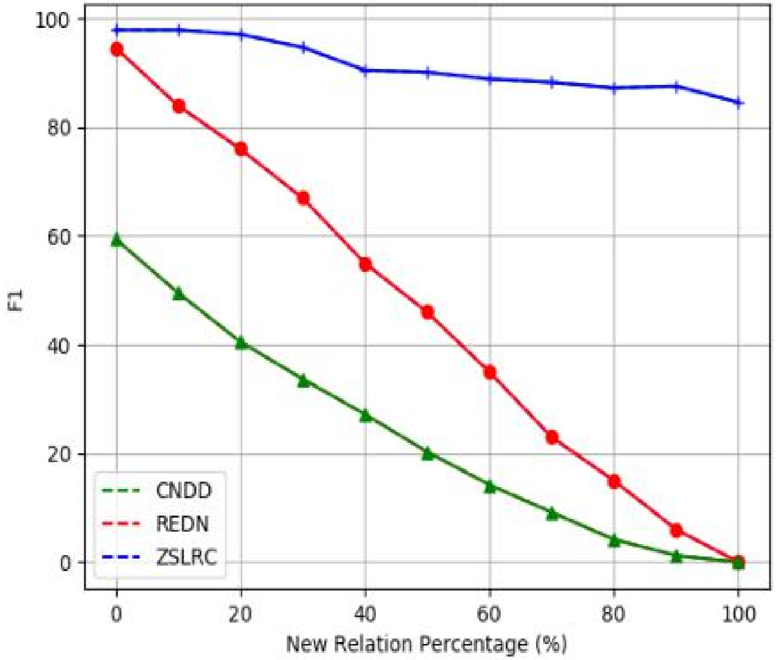


Figure 5: F1-score of ZSLRC when different proportions of new relations appear in NYT dataset.

Experiments

Table 3: Ablation Results on NYT dataset (Accuracy%)

	10%	30%	50%	70%	90%
ZSLRC(HE)	88.94	70.57	52.12	33.87	15.48
ZSLRC(KE)	93.12	82.22	71.00	60.47	49.07
ZSLRC(SIE)	93.86	85.14	81.91	78.79	72.57
ZSLRC(WSIE)	**96.64**	**94.46**	**92.14**	**91.82**	**89.3**

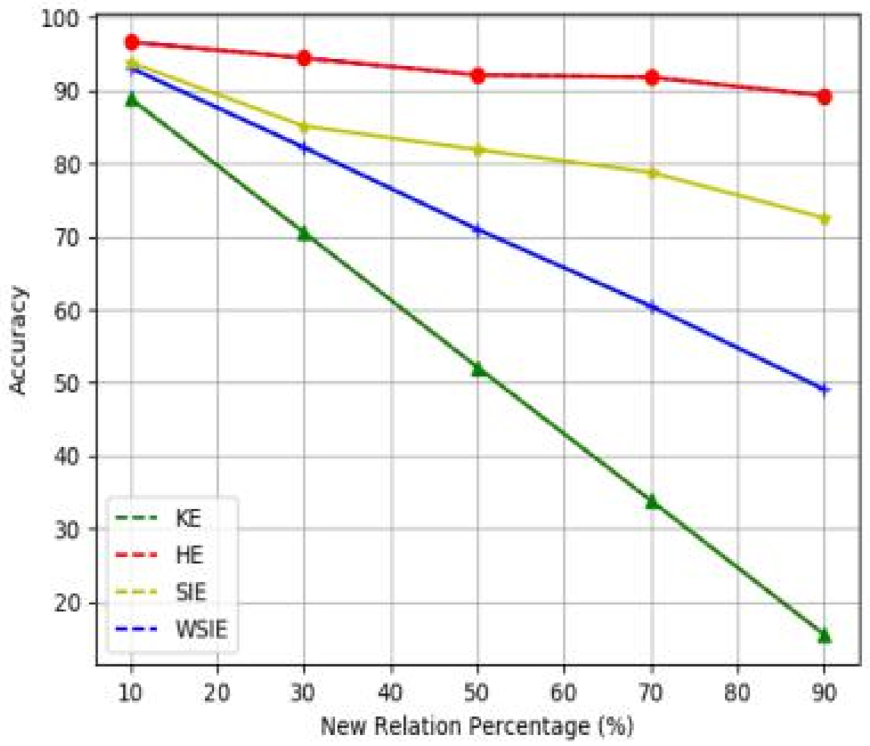


Figure 6: Ablation study of ZSLRC on NYT dataset.

Experiments

Table 4: Results of Accuracy Comparison Among Models (%)

Model	5 way 1 shot	5 way 5 shot	5 way 10 shot	10 way 1 shot	10 way 5 shot	10 way 10 shot
Meta Network*	64.46 ± 0.54	80.57 ± 0.48	-	53.96 ± 0.56	69.23 ± 0.52	-
GNN*	66.23 ± 0.75	81.28 ± 0.62	-	46.27 ± 0.80	64.02 ± 0.77	-
SNAIL*	67.29 ± 0.26	79.40 ± 0.22	-	53.28 ± 0.27	68.33 ± 0.25	-
Proto(CNN)	73.62 ± 0.20	85.78 ± 0.16	88.45 ± 0.10	60.96 ± 0.22	75.38 ± 0.19	78.71 ± 0.11
Proto-HATT(CNN)	74.68 ± 0.18	86.73 ± 0.12	89.64 ± 0.12	61.61 ± 0.16	77.04 ± 0.12	79.99 ± 0.11
Proto-CATT(CNN)	-	87.48 ± 0.12	89.28 ±0.08	-	77.46 ± 0.13	80.39 ± 0.14
ZSLRC(CNN)	**75.83±0.17**	**87.84±0.12**	**89.67±0.12**	**63.54±0.14**	**77.64±0.11**	**80.69±0.10**

Note that to fairly compare the performance of each model, we only compare the models with the same 50-dimension GloVe embedding and CNN encoders of the same parameters. Better results can be achieved through the BERT encoder.

Table 5: Ablation Results on FewRel dataset (%).

Model	5 way 1 shot	5 way 5 shot	5 way 10 shot	10 way 1 shot	10 way 5 shot	10 way 10 shot
Proto(CNN)	73.62 ± 0.20	85.57 ± 0.14	88.17 ± 0.10	62.22 ± 0.32	75.01 ± 0.16	78.50 ± 0.11
ZSLRC(HE)	75.66 ± 0.14	86.55 ± 0.13	88.98 ± 0.10	63.28 ± 0.20	76.58 ± 0.06	79.93 ± 0.05
ZSLRC(KE)	74.57 ± 0.08	86.70 ± 0.17	89.09 ± 0.11	62.39 ± 0.12	76.99 ±0.20	80.06 ± 0.09
ZSLRC(SIE)	75.56 ± 0.12	87.34 ± 0.14	89.17 ± 0.13	63.02 ± 0.15	77.16 ± 0.12	80.34 ± 0.10
ZSLRC(WSIE)	**75.83±0.17**	**87.84±0.12**	**89.67±0.12**	**63.54±0.14**	**77.64±0.11**	**80.69±0.10**
ZSLRC(WSIEA)	75.58 ± 0.15	87.16 ± 0.16	89.17 ± 0.15	62.85 ± 0.18	76.71 ± 0.14	80.18 ±0.11

Experiments

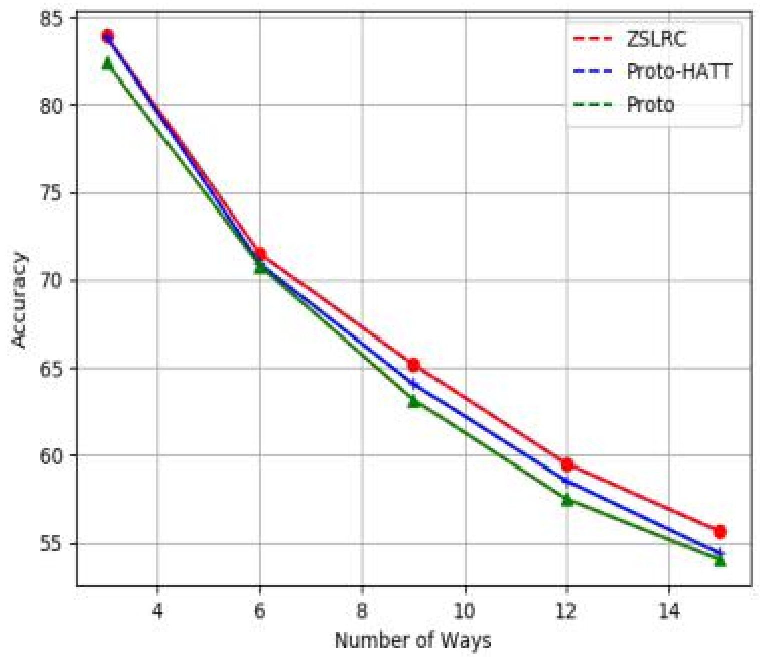


Figure 7: Accuracy of our proposed model in different N-way One-shot tasks.

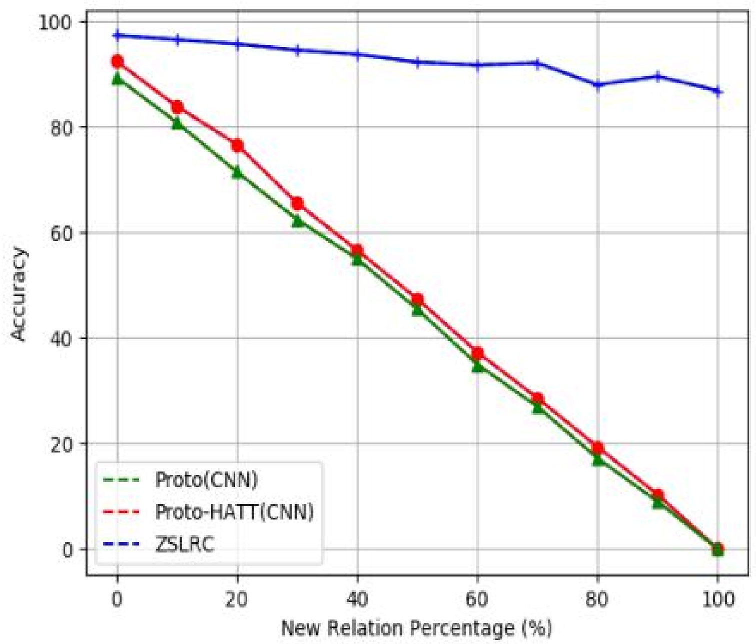


Figure 8: Accuracy of ZSLRC when different proportions of new relations appear in re-splitted FewRel dataset.

Thank you!