Modeling Fine-grained Information via Knowledge-aware Hierarchical Graph for Zero-shot Entity Retrieval

Taiqiang Wu*
wtq20@mails.tsinghua.edu.cn
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China

Xingyu Bai*
bxy20@mails.tsinghua.edu.cn
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China

Weigang Guo
jimwgguo@tencent.com
Tencent
Shenzhen, China

Weijie Liu
jagerliu@tencent.com
Tencent
Shenzhen, China

Siheng Li
lisiheng21@mails.tsinghua.edu.cn
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China

Yujiu Yang†
yang.yujiu@sz.tsinghua.edu.cn
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China

（WSDM-2023)

code： None

gesis
Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S

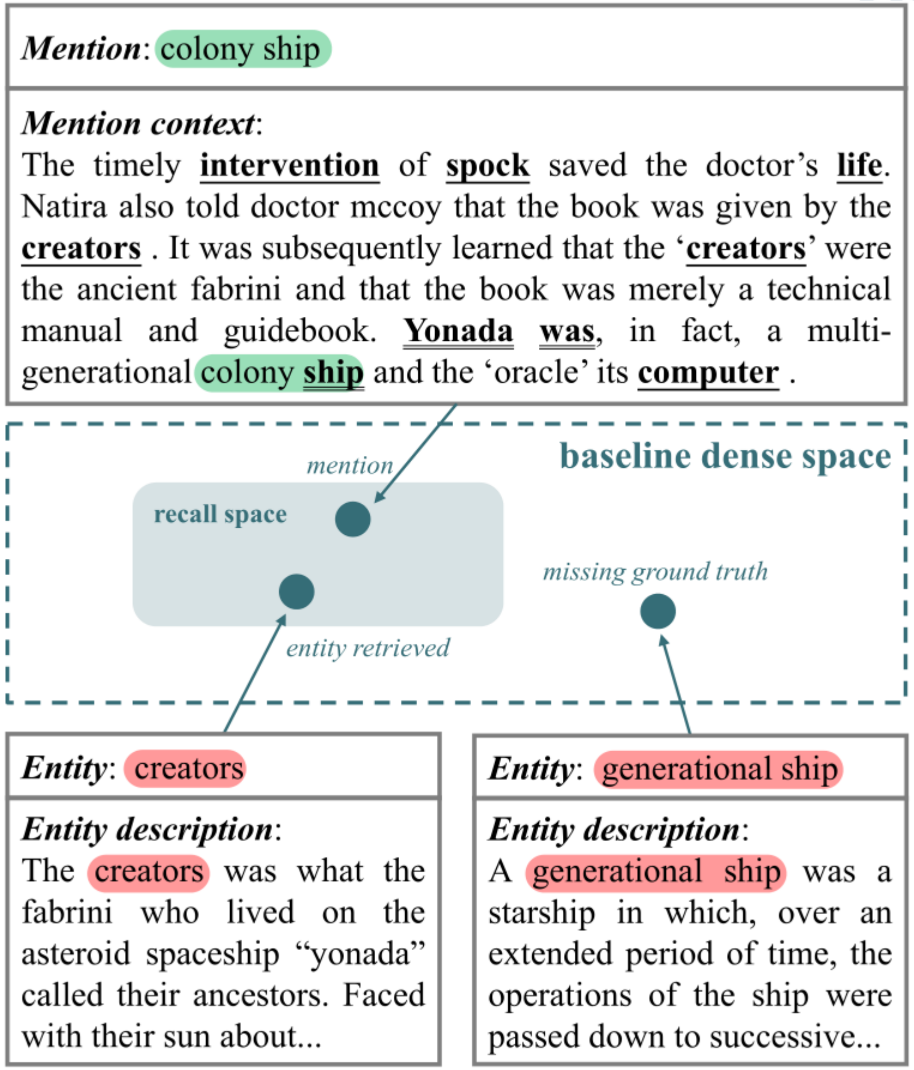
Reported by Zhaoze Gao

1. Introduction

2. Approach

3. Experiments

Introduction



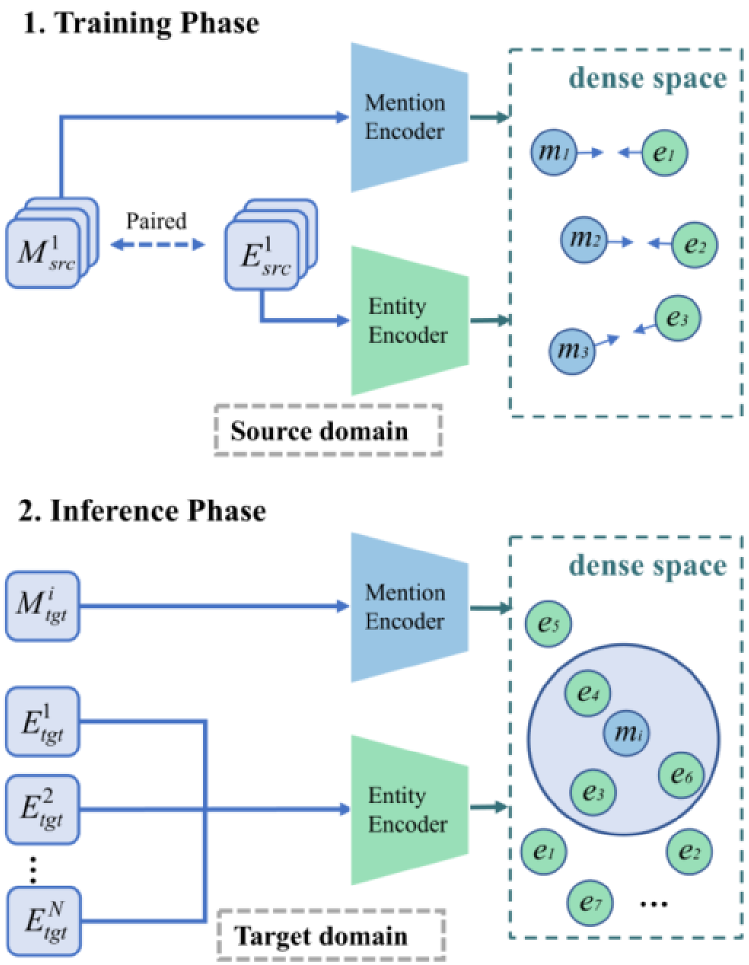


Figure 2: Overview of our GER framework.

Approach

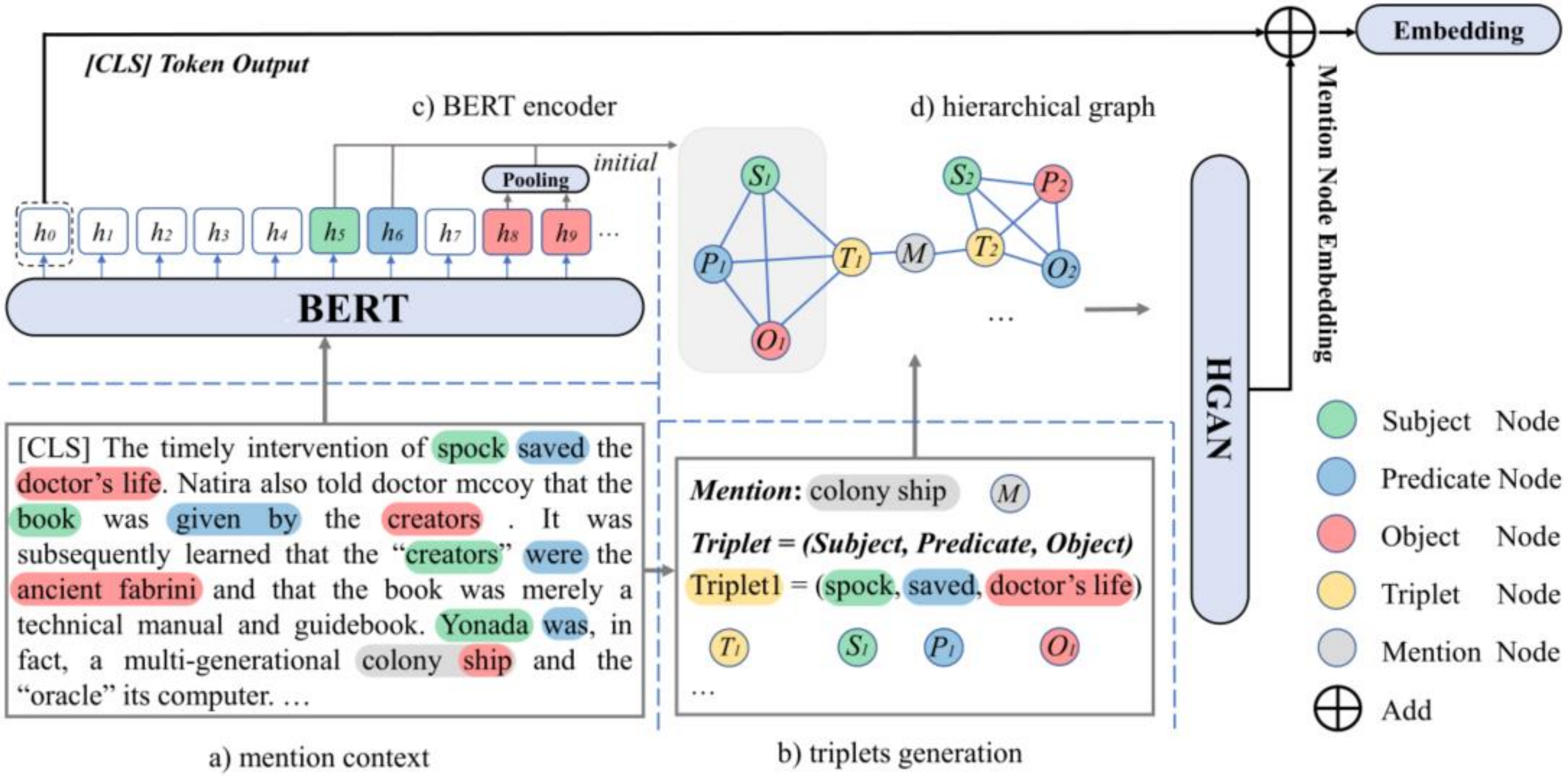
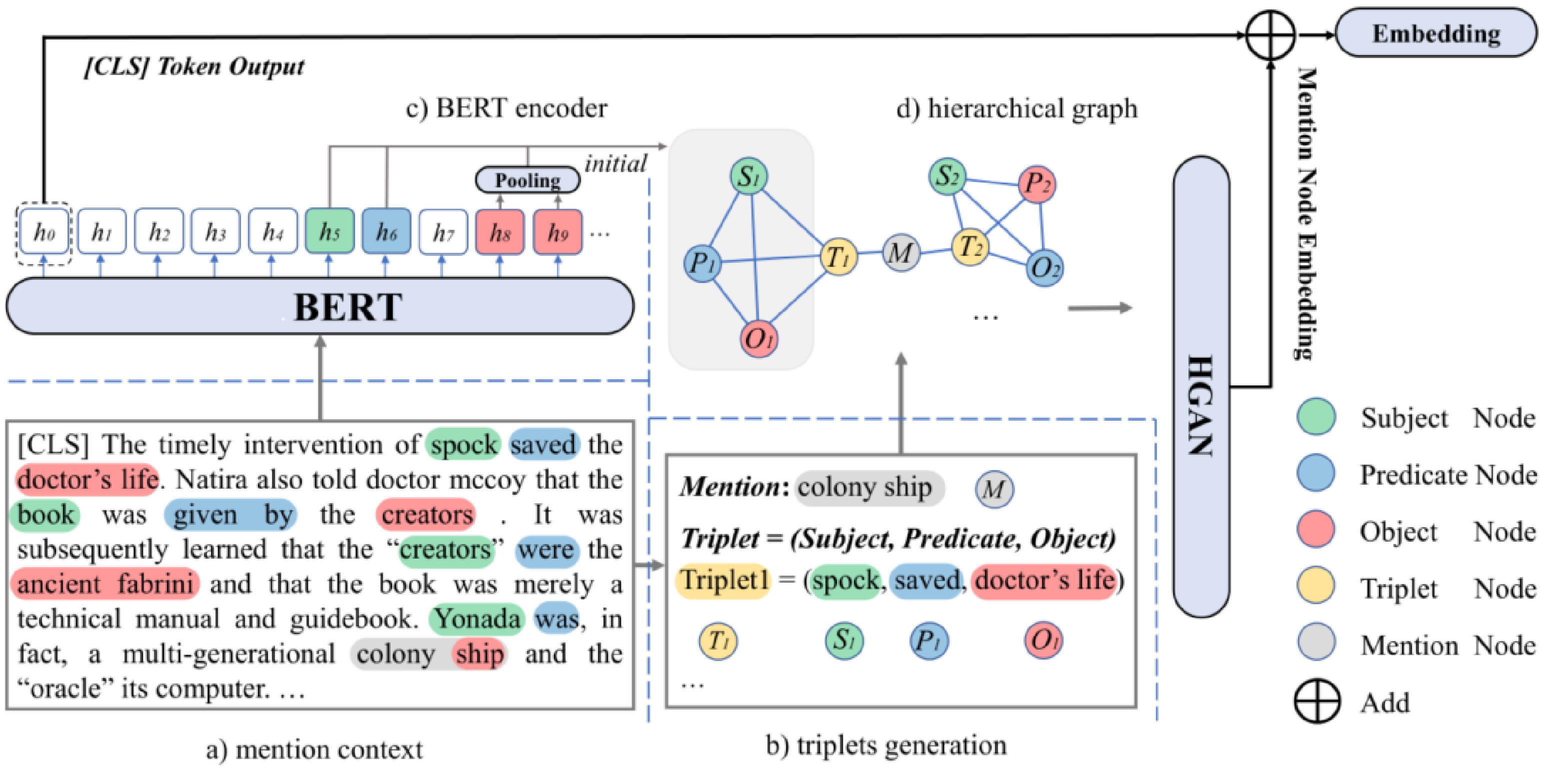


Figure 3: Overview of mention encoder in GER. For the given mention representation (shown in part a), we extract the triplets (shown in part b, green for the subject, blue for the predicate, and red for the object) as knowledge units. To avoid the graph bottleneck [2], we add a triplet node (in yellow) between the mention/entity node and each triplet, and thus build the hierarchical graph (shown in part d).

Approach



$$Y_m = T_m(\texttt{[CLS]}\ c_l\ \texttt{[MS]}\ m\ \texttt{[ME]}\ c_r) \quad (1)$$

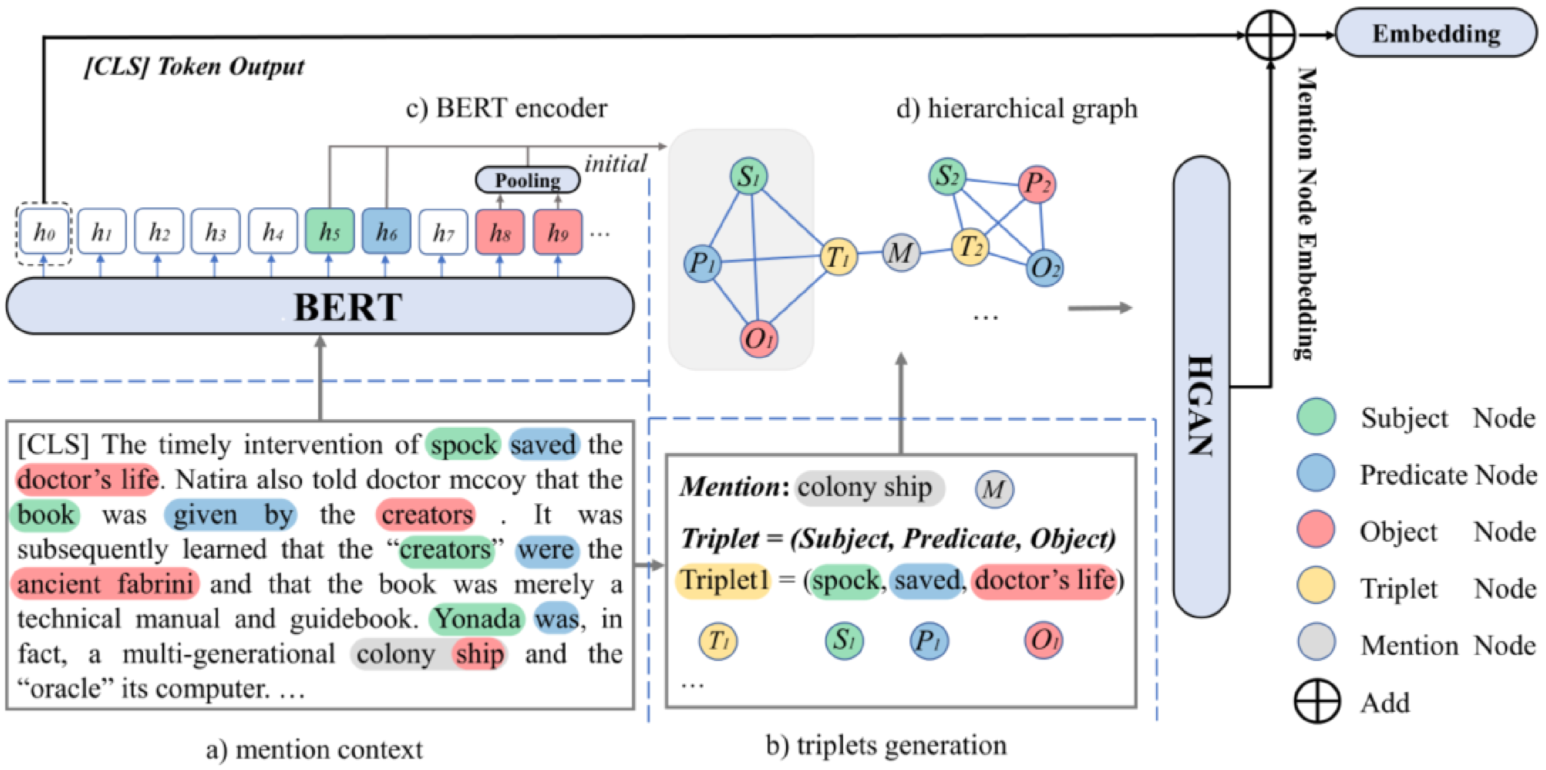
$$Y_e = T_e(\texttt{[CLS]}\ e\ \texttt{[ENT]}\ d) \quad (2)$$

$$\mathbf{h}_m^0 = red(Y_m[p_{start} : p_{end}]) \quad (3)$$

$$\mathbf{h}_t^0 = \left[\mathbf{h}_s^0 \| \mathbf{h}_p^0 \| \mathbf{h}_o^0\right] \mathbf{W}^{triple} \quad (4)$$

$$\mathbf{h}_i^{(l)} = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in N_i\cup\{i\}} \alpha_{ij}^k \mathbf{h}_j^{(l-1)} \mathbf{W}^k\right) \quad (5)$$

Approach



$$\alpha_{ij}^{k} = \frac{\exp(e_{ij}^{k})}{\sum_{j'\in\mathcal{N}_i\cup\{i\}}\exp(e_{ij'}^{k})} \tag{6}$$

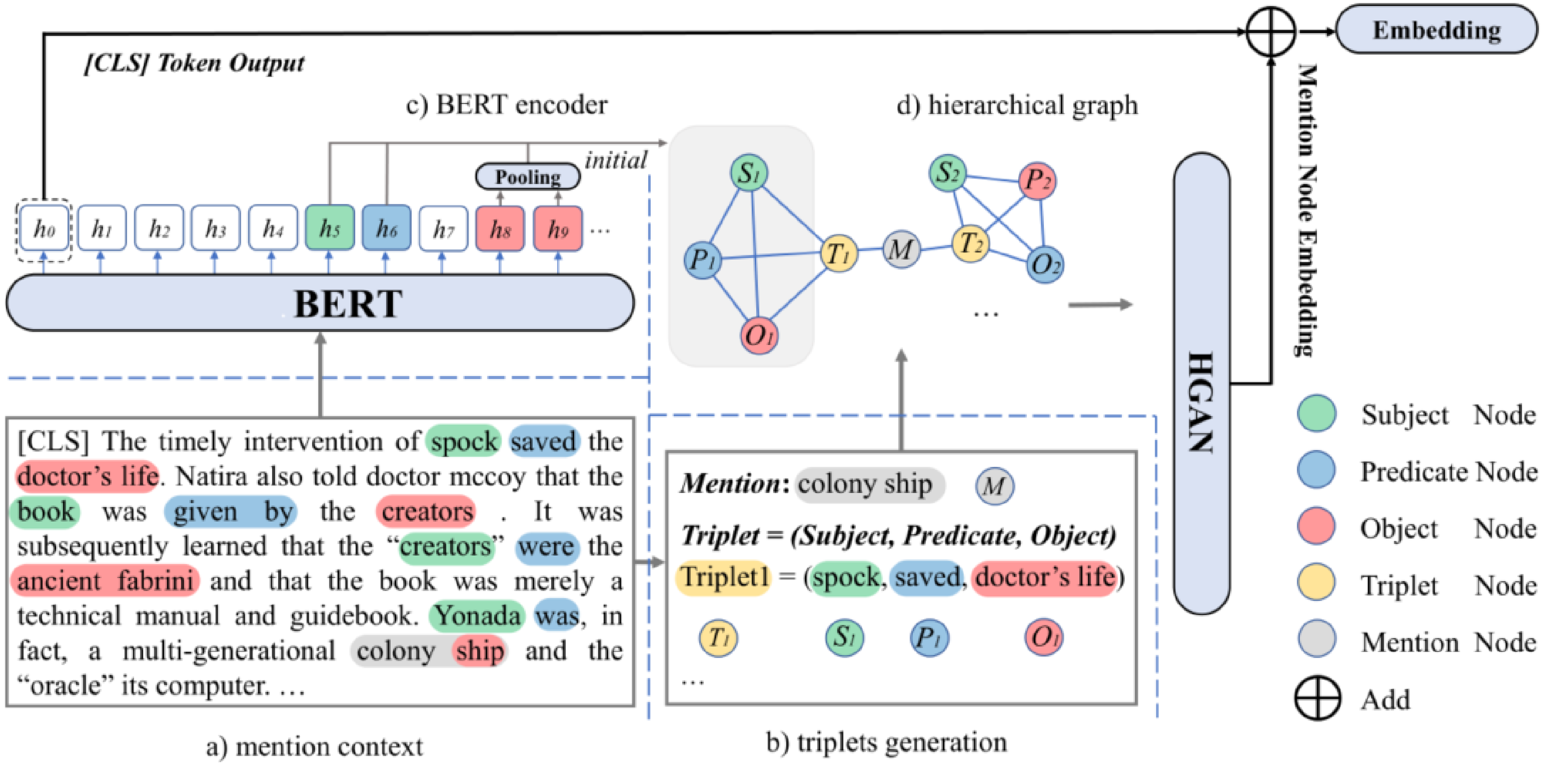
$$e_{ij'}^{k} = \mathtt{LeakyReLU}([\mathbf{h}_i\mathbf{W}^k \| \mathbf{h}_{j'}\mathbf{W}^k]\mathbf{a}) \tag{7}$$

$$v^{sen} = Y_m[0] \tag{8}$$

$$v^{graph} = h_m^L \tag{9}$$

$$v = v^{sen} + \lambda v^{graph} \tag{10}$$

Approach



$$\mathcal{L}(m_i, e_i) = \mathcal{L}_1(m_i, e_i) + \mathcal{L}_2(m_i, e_i) \tag{11}$$

$$\mathcal{L}_1(m_i, e_i) = -s(m_i, e_i) + \log \sum_{j=1}^{bsz} \exp(s(m_i, e_j)) \tag{12}$$

$$\mathcal{L}_2(m_i, e_i) = -s(m_i, e_i) + \log \sum_{i=1}^{bsz} \exp(s(m_i, e_j)) \tag{13}$$

$$s(m_i, e_i) = v_{m_i} \cdot v_{e_i}^T \tag{14}$$

Experiments

KB	Dataset	Usage	Samples Num	Entity Num
Wiki-pedia	AIDA	Train	18,317	5,903,530
		Valid	4,763	
	WNED-CWEB	Test	10,392	
	AQUAINT	Test	678	
Wikia	ZESHEL	Train	49,275	332,632
		Valid	10,000	89,549
		Test	10,000	70,140

Table 1: Statistics of entity retrieval datasets and knowledge base, samples num means the size of paired mentions and entities. For each KB, we use the corresponding train dataset (e.g., AIDA train set) to optimize our GER framework, and report the recall results on test dataset (e.g., WNED-CWEB).

Experiments

Method	R@1	R@4	R@8	R@16	R@32	R@50	R@64
BM25 [17]†	-	-	-	-	-	-	69.13
BLINK [30]†	-	-	-	-	-	-	82.06
Partalidou et al. [19]†	-	-	-	-	-	84.28	-
ARBORESCENCE [1]†	-	-	-	-	-	-	85.11
BLINK [30]*	38.01	62.08	69.19	75.39	80.03	82.69	83.98
BERT Mean Pooling	33.65	57.74	65.17	71.38	75.85	78.66	80.14
BERT Max Pooling	36.94	60.42	68.34	73.83	78.40	81.09	82.65
BLINK + BERT Mean Pooling	34.12	58.41	66.19	72.24	76.93	79.79	81.16
BLINK + BERT Max Pooling	38.45	63.46	70.68	76.72	81.11	83.63	84.83
GER (ours)	**42.86**	**66.48**	**73.00**	**78.11**	**82.15**	**84.41**	**85.65**

Table 2: *Recall@K* (R@K) results on the test set of ZESHEL dataset, which is the average of 5 runs with different random seeds. * notes for the results we reproduce. † notes for the results taken from their papers. Best results are shown in bold. GER outperforms all baselines significantly with paired t-test at $p < 0.05$ level considering R@64.

Experiments

Method	WNED-CWEB			AQUAINT		
	R@10	**R@30**	**R@128**	**R@10**	**R@30**	**R@128**
BLINK*	80.16	84.48	89.22	93.95	96.76	98.23
BERT Mean Pooling	79.87	84.79	89.35	94.54	96.90	98.23
BERT Max Pooling	77.62	83.56	88.57	93.07	95.87	97.94
BLINK + BERT Mean Pooling	80.13	84.33	88.81	94.84	96.31	98.23
BLINK + BERT Max Pooling	78.75	84.16	88.81	93.22	96.02	97.35
GER (ours)	**80.79**	**85.34**	**90.13**	**95.28**	**97.05**	**98.82**

Table 3: *Recall@K* (R@K) on dataset WNED-CWEB and AQUAINT. The experiments are all under the zero-shot settings. that entities are only defined by textual description and the entities in test set are unseen during training.

Experiments

Sentence-level	Word-level	R@1	R@8	R@32	R@64
BERT	-	38.01	69.19	80.03	83.98
-	HGAN	37.37	63.77	73.19	77.29
BERT	Node Mean	37.29	69.62	80.15	83.88
BERT	GAT	39.23	70.07	80.14	84.09
BERT	HGAN	42.86	73.00	82.15	85.65

Table 4: The ablation study results of our GER (BERT+HGAN) on ZESHEL dataset.

Experiments

Mention Encoder	Entity Encoder	R@1	R@8	R@32	R@64
BERT	BERT	38.01	69.19	80.03	83.98
BERT+HGAN	BERT	38.16	69.41	80.04	83.92
BERT	BERT+HGAN	39.18	68.56	78.70	82.65
BERT+HGAN	BERT+HGAN	42.86	73.00	82.15	85.65

Table 5: The ablation study results of the dual-encoder architecture. (BERT, BERT) is the baseline BLINK while (BERT+HGAN, BERT+HGAN) is our proposed GER.

Experiments

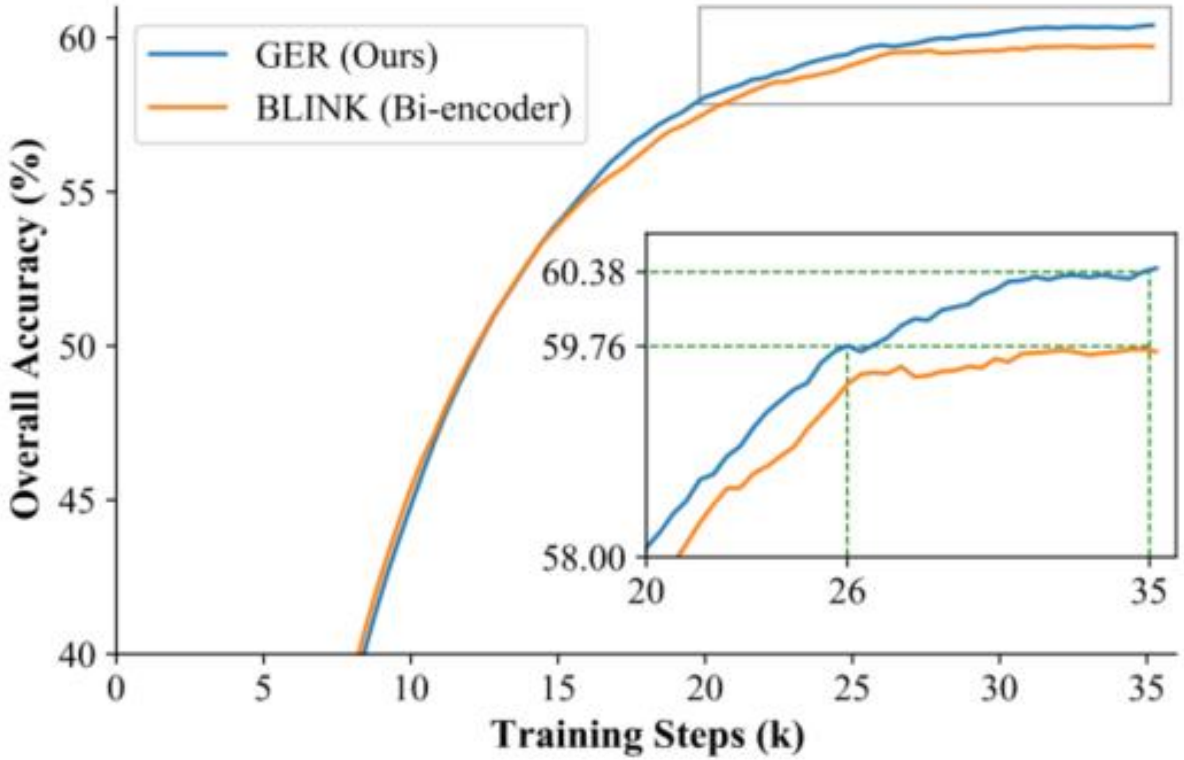


Figure 4: Comparison of overall accuracy for BLINK and GER.

Experiments

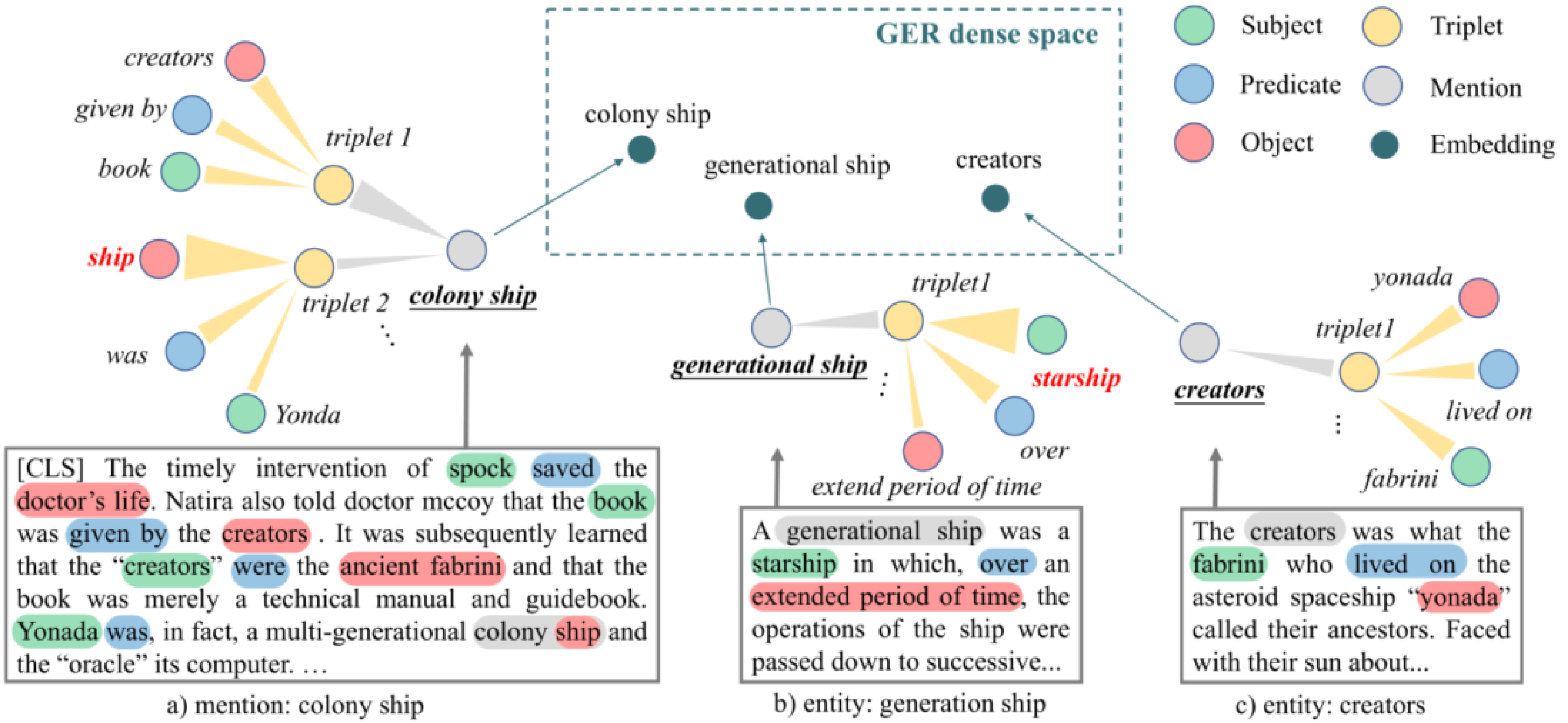


Figure 5: The corresponding embedding in dense space and part of node attention between graph nodes for mention *colony ship*, ground truth entity *generation ship* and entity *creators*. In the graph, we visualize the attention of Mention/Entity-Triplet edges (in grey) and Triplet-SPO edges (in yellow), where *thicker edges* mean *higher* attention scores.

Experiments

Attention Ranking		[0,32)	[32,64)	[64,96)	[96,128)
BLINK	Total	685	1191	2765	5359
	Recall@64	86.13	85.81	84.45	83.50
GER	Total	742	1315	2798	5145
	Recall@64	86.12	86.08	86.78	84.98

Table 6: Attention distributions for ZESHEL test set.

Thank you！