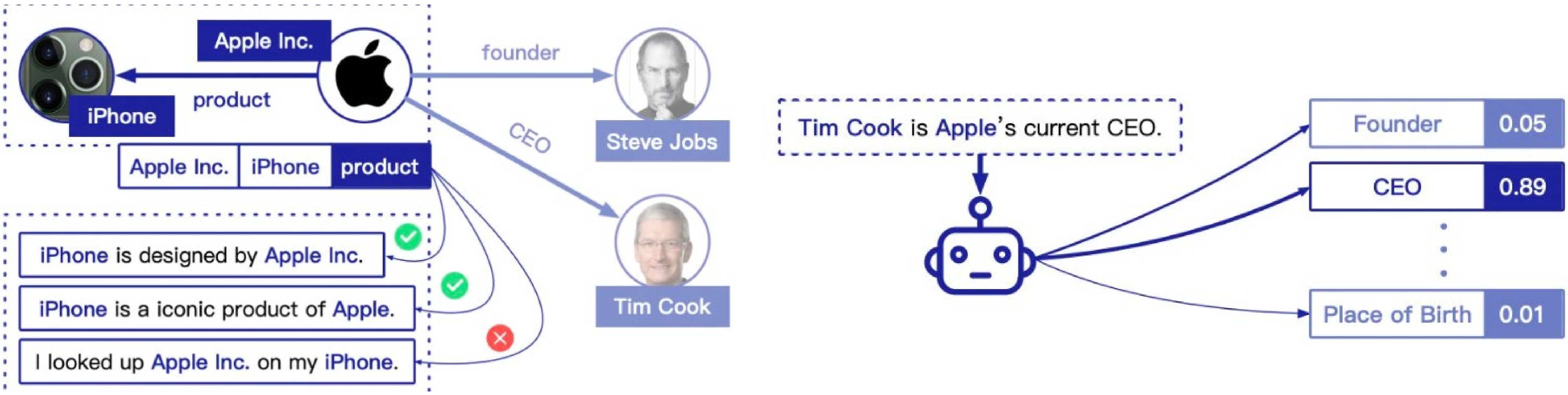
Relation Extraction Task

伊鑫

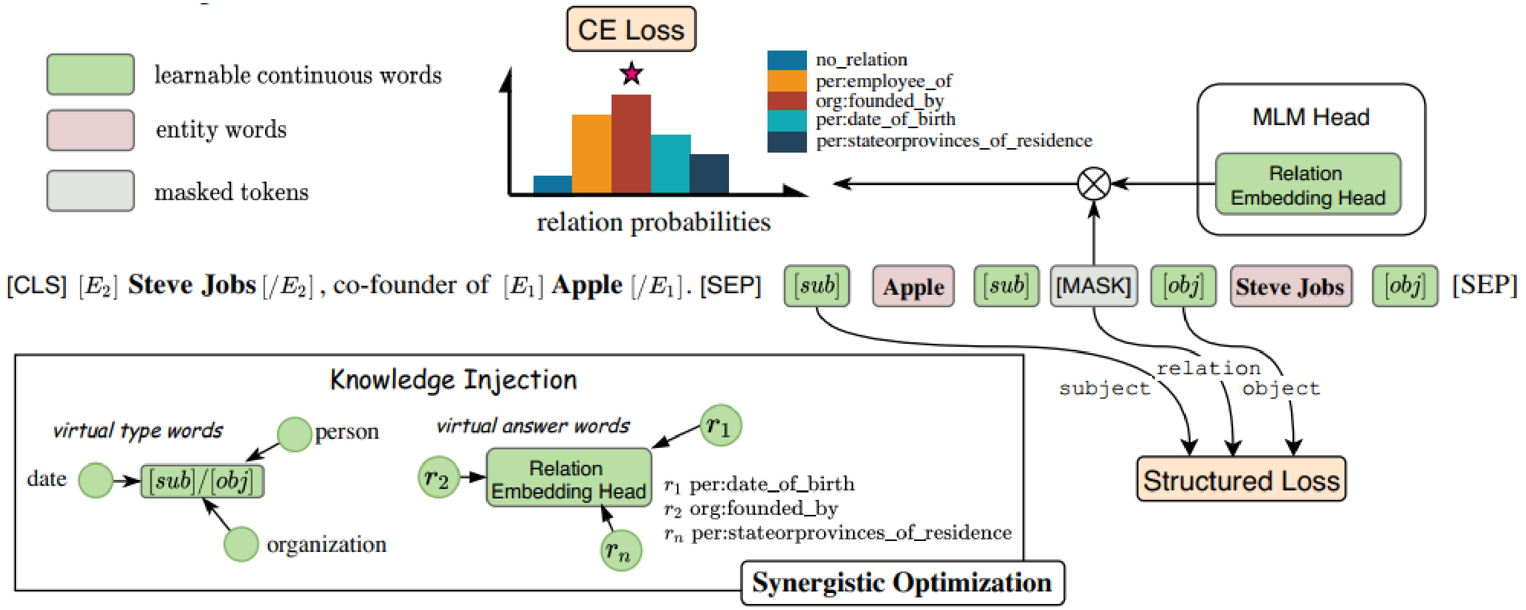
52265901030@stu.ecnu.edu.cn

2022.10.25

Relation extraction introduction



virtual type words and answer words



$$\mathcal{J}_{\text{structured}} = -\log \sigma(\gamma - d_r(\mathbf{s}, \mathbf{o}))$$

$$-\sum_{i=1}^{n} \frac{1}{n} \log \sigma(d_r(\mathbf{s}_i', \mathbf{o}_i') - \gamma),$$

$$d_r(\mathbf{s}, \mathbf{o}) = \|\mathbf{s} + \mathbf{r} - \mathbf{o}\|_2,$$

Relation Labels	C_{sub}	C_{obj}
per:country_of_birth	person	country
per:data_of_death	person	data
per:schools_attended	person	organization
org:alternate_names	organization	organization
org:city_of_headquarters	organization	city
org:number_of_employees/members	organization	number

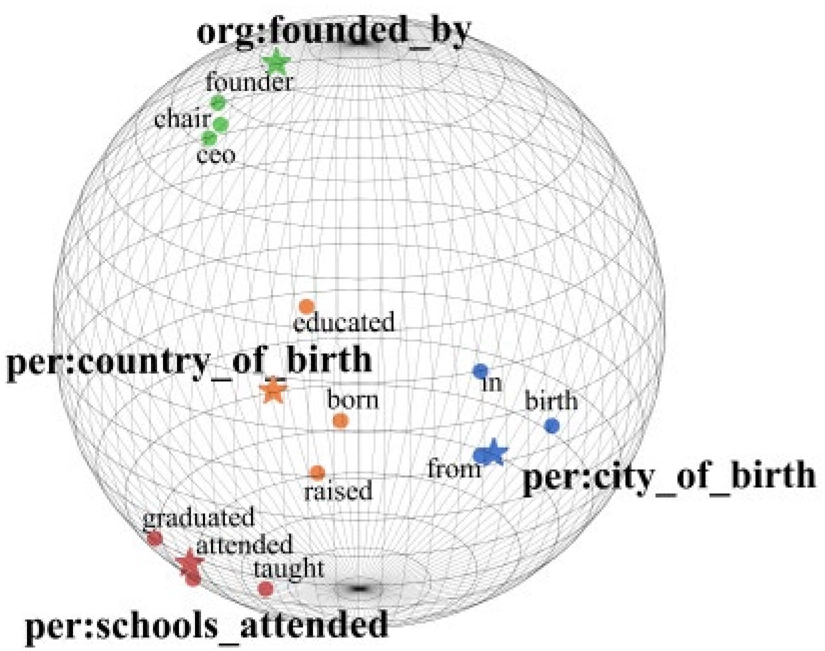
C_r (Disassembled Relation Prepared for Virtual Answer Words)
{"country", "of", "birth" }
{"data", "of", "death" }
{"school","attended"}
{"alternate", "names" }
{"city", "of", "headquarters" }
{"number", "of", "employees", "members" }

KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction (WWW 2022)

Experiments

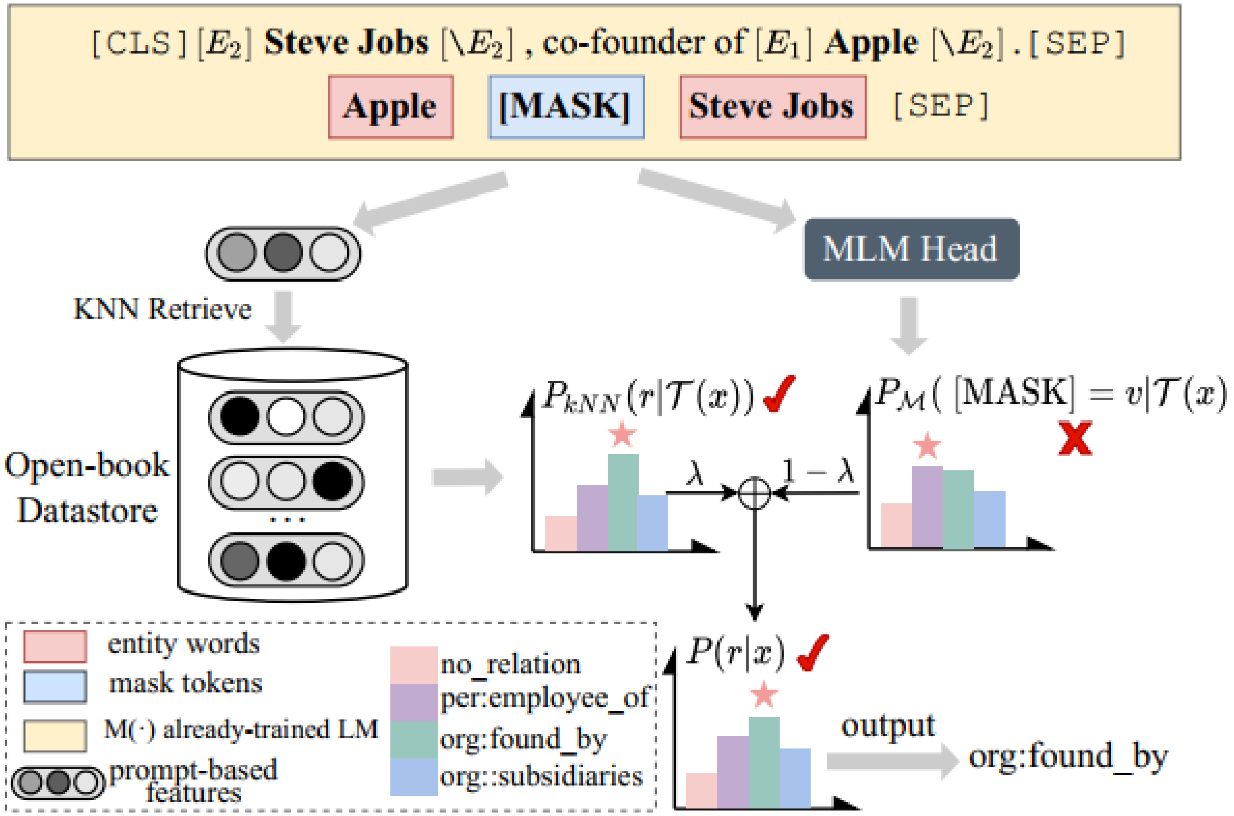
Standard Supervised Setting						
Methods	Extra Data	SemEval	DialogRE†	TACRED	TACRED-Revisit	Re-TACRED
Fine-tuning pre-trained models						
Fine-tuning-[RoBERTa]	w/o	87.6	57.3	68.7	76.0	84.9
SpanBERT [30]	w/	-	-	70.8	78.0	85.3
KnowBERT [38]	w/	89.1	-	71.5	79.3	89.1
LUKE [52]	w/	-	-	**72.7**	80.6	-
MTB [3]	w/	89.5	-	70.1	-	-
GDPNet [51]	w/o	-	64.9	71.5	79.3	-
Dual [2]	w/o	-	67.3	-	-	-
Prompt-tuning pre-trained models						
PTR-[RoBERTa] [22]	w/o	89.9	63.2	72.4	81.4	90.9
KnowPrompt-[RoBERTa]	w/o	**90.2** (+0.3)	**68.6** (+5.4)	72.4 (-0.3)	**82.4** (+1.0)	**91.3** (+0.4)

Low-Resource Setting							
Split	Methods	SemEval	DialogRE†	TACRED	TACRED-Revisit	Re-TACRED	Average
K=8	Fine-tuning	41.3	29.8	12.2	13.5	28.5	25.1
	GDPNet	42.0	28.6	11.8	12.3	29.0	24.7
	PTR	70.5	35.5	28.1	28.7	51.5	42.9
	KnowPrompt	**74.3** (+33.0)	**43.8** (+14.0)	**32.0** (+19.8)	**32.1** (+18.6)	**55.3** (+26.8)	**47.5** (+22.4)
K=16	Fine-tuning	65.2	40.8	21.5	22.3	49.5	39.9
	GDPNet	67.5	42.5	22.5	23.8	50.0	41.3
	PTR	81.3	43.5	30.7	31.4	56.2	48.6
	KnowPrompt	**82.9** (+17.7)	**50.8** (+10.0)	**35.4** (+13.9)	**33.1** (+10.8)	**63.3** (+13.8)	**53.1** (+13.2)
K=32	Fine-tuning	80.1	49.7	28.0	28.2	56.0	48.4
	GDPNet	81.2	50.2	28.8	29.1	56.5	49.2
	PTR	84.2	49.5	32.1	32.4	62.1	52.1
	KnowPrompt	**84.8** (+4.7)	**55.3** (+3.6)	**36.5** (+8.5)	**34.7** (+6.5)	**65.0** (+9.0)	**55.3** (+6.9)



A 3D visualization of several relation representations (virtual answer words) optimized in KnowPrompt on TACRED-Revisit dataset using t-SNE and normalization.

Retrieval Open-book



training set to construct the datastore

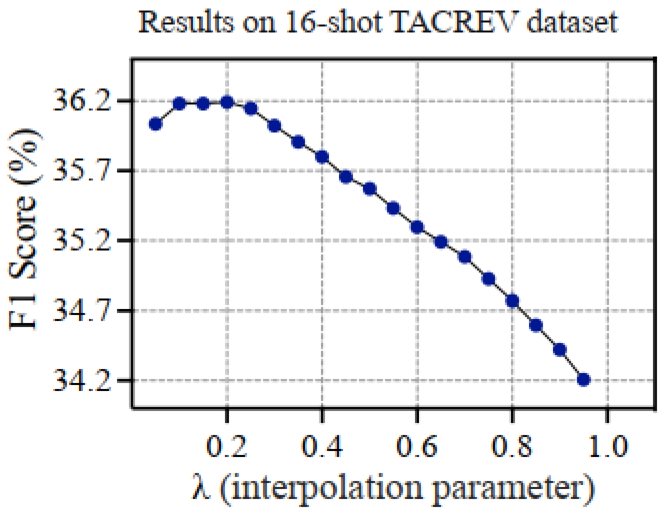
retrieve the k-nearest neighbors

$$P_{\mathrm{kNN}}(r \mid \mathcal{T}(x)) \propto \sum_{(c_i, r_i) \in \mathcal{N}} \mathbb{1}_{r=r_i} \exp\left(-d\left(h_x, h_{c_i}\right)\right)$$

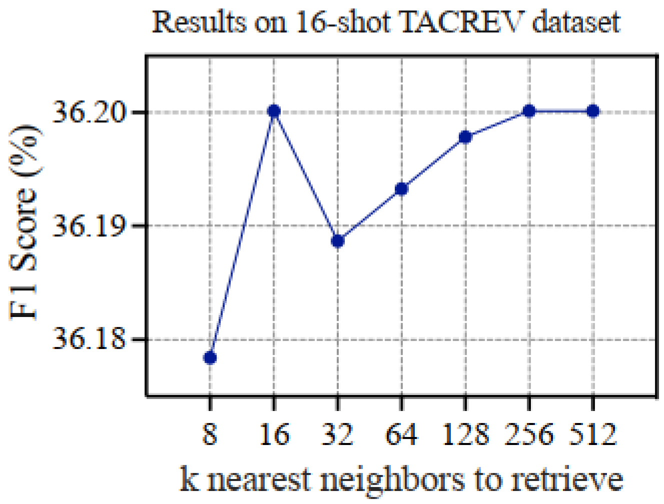
$$P(r \mid x) = \lambda P_{\mathrm{kNN}}(r \mid \mathcal{T}(x)) + (1-\lambda) P_{\mathcal{M}}(\text{[MASK]} = v|\mathcal{T}(x))$$

Relation Extraction as Open-book Examination: Retrieval-enhanced Prompt Tuning (SIGIR 2022)

Experiments



(a) λ varies.

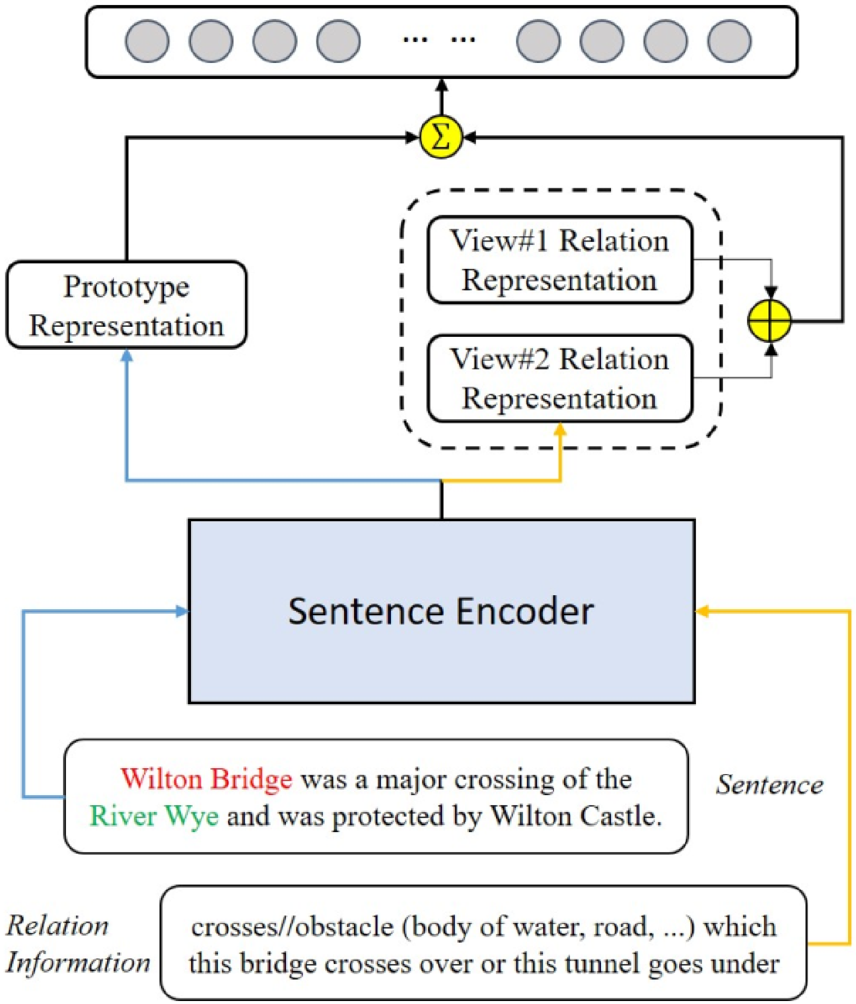


(b) k varies.

Standard Supervised Setting					
Methods	SemEval	DialogRE	TACRED	TACREV	Re-TACRED
Fine-tuning pre-trained models					
Fine-tuning	87.6	57.3	68.7	76.0	84.9
SpanBERT [22]	-	-	70.8	78.0	85.3
KnowBERT [30]	89.1	-	71.5	79.3	89.1
LUKE [38]	-	-	**72.7**	80.6	-
MTB [4]	89.5	-	70.1	-	-
GDPNet [37]	-	64.9	71.5	79.3	-
Dual [3]	-	67.3	-	-	-
Prompt-tuning pre-trained models					
PTR [13]	89.9	63.2	72.4	81.4	90.9
KnowPrompt	90.2	68.6	72.4	82.4	91.3
RetrievalRE	**90.4**	**69.4**	**72.7**	**82.7**	**91.5**

Source	Model	SemEval			TACRED			TACREV		
		K=1	K=5	K=16	K=1	K=5	K=16	K=1	K=5	K=16
None	Fine-tuning	18.5 (± 1.4)	41.5 (± 2.3)	66.1 (± 0.4)	7.6 (± 3.0)	16.6 (± 2.1)	26.8 (± 1.8)	7.2 (± 1.4)	16.3 (± 2.1)	25.8 (± 1.2)
	GDPNet	10.3 (± 2.5)	42.7 (± 2.0)	67.5 (± 0.8)	4.2 (± 3.8)	15.5 (± 2.3)	28.0 (± 1.8)	5.1 (± 2.4)	17.8 (± 2.4)	26.4 (± 1.2)
	PTR	14.7 (± 1.1)	53.9 (± 1.9)	80.6 (± 1.2)	8.6 (± 2.5)	24.9 (± 3.1)	30.7 (± 2.0)	9.4 (± 0.7)	26.9 (± 1.5)	31.4 (± 0.3)
	KnowPrompt	28.6 (± 6.2)	66.1 (± 8.6)	80.9 (± 1.6)	17.6 (± 1.8)	28.8 (± 2.0)	34.7 (± 1.8)	17.8 (± 2.2)	30.4 (± 0.5)	33.2 (± 1.4)
	RetrievalRE	**33.3** (± 1.6)	**69.7** (± 1.7)	**81.8** (± 1.0)	**19.5** (± 1.5)	**30.7** (± 1.7)	**36.1** (± 1.2)	**18.7** (± 1.8)	**30.6** (± 0.2)	**35.3** (± 0.3)

Prototype representation



- support set S and query set Q
- View #1 embedding of the "[CLS]“ token
- View #2 the average of the embeddings of all tokens
- Prototype Represstentation: $[h_{entity1}; h_{entity2}]$
- vector dot product way to calculate the distance between the query instance Q and each class prototype
- CE loss

Experiments

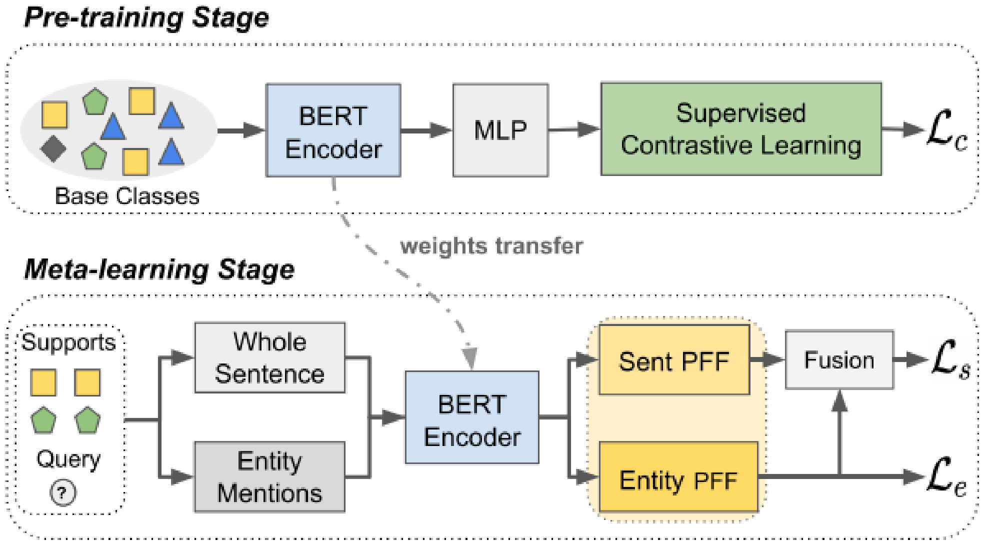
Encoder	Model	5-w-1-s	5-w-5-s	10-w-1-s	10-w-5-s
CNN	Proto-HATT	72.65 / 74.52	86.15 / 88.40	60.13 / 62.38	76.20 / 80.45
	MLMAN	75.01 / ——	87.09 / 90.12	62.48 / ——	77.50 / 83.05
BERT	BERT-PAIR	85.66 / 88.32	89.48 / 93.22	76.84 / 80.63	81.76 / 87.02
	Proto-BERT*	84.77 / 89.33	89.54 / 94.13	76.85 / 83.41	83.42 / 90.25
	REGRAB	87.95 / 90.30	92.54 / 94.25	80.26 / 84.09	86.72 / 89.93
	TD-proto	—— / 84.76	—— / 92.38	—— / 74.32	—— / 85.92
	CTEG	84.72 / 88.11	92.52 / 95.25	76.01 / 81.29	84.89 / 91.33
	ConceptFERE	—— / 89.21	—— / 90.34	—— / 75.72	—— / 81.82
	HCRP (BERT)	90.90 / 93.76	93.22 / 95.66	84.11 / 89.95	87.79 / 92.10
	Ours (BERT)	91.29 / **94.42**	94.05 / **96.37**	86.09 / **90.73**	89.68 / **93.47**
	MTB	—— / 91.10	—— / 95.40	—— / 84.30	—— / 91.80
	CP	—— / 95.10	—— / 97.10	—— / 91.20	—— / 94.70
	MapRE	—— / 95.73	—— / 97.84	—— / 93.18	—— / 95.64
	HCRP (CP)	94.10 / 96.42	96.05 / **97.96**	89.13 / 93.97	93.10 / **96.46**
	Ours (CP)	96.21 / **96.63**	97.07 / 97.93	93.38 / **94.94**	95.11 / 96.39
	Δ	+5.09	+2.24	+7.32	+3.22
	Δ (CP)	+1.53	+0.83	+3.74	+1.69

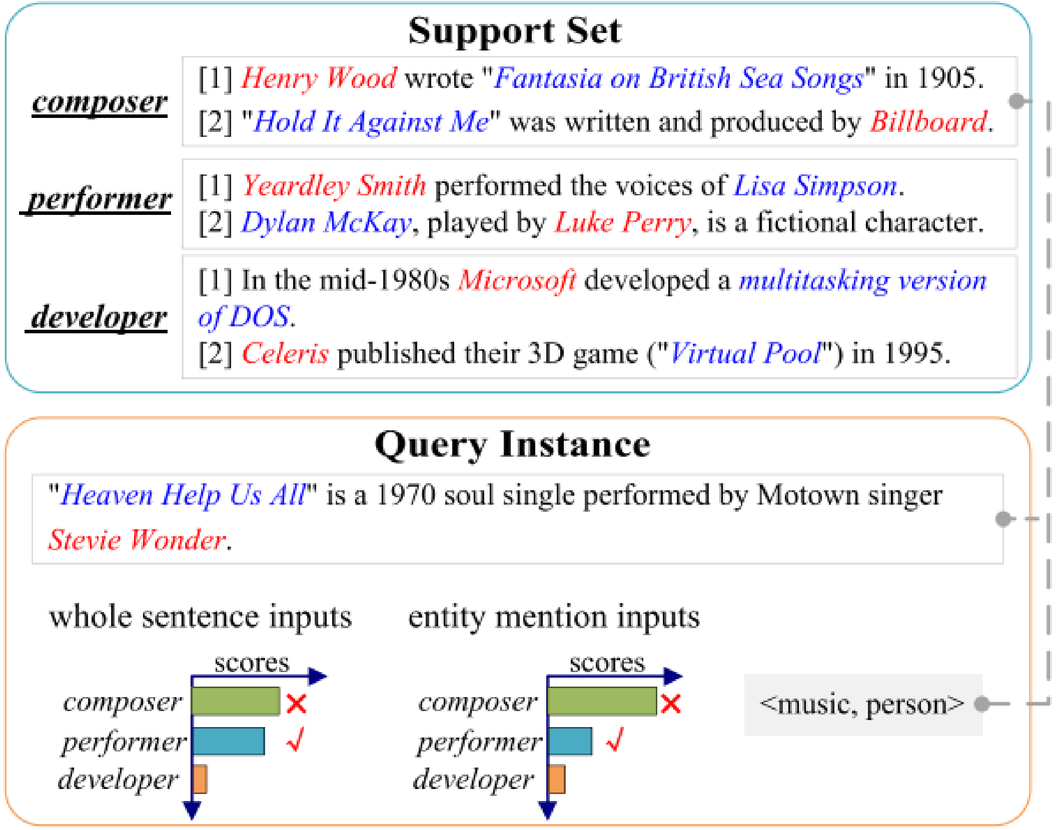
Experimental results of FSRE on FewRel 1.0 validation / test set,

Model	5-w-1-s	10-w-1-s
Ours	91.29	86.09
w/o relation info.	84.77	76.85
w/ concat	79.16	65.12
w/ linear layer		
view#1	89.04	80.29
view#2	89.39	80.14

Ablation Study on validation set of FewRel

Unbiased Representations





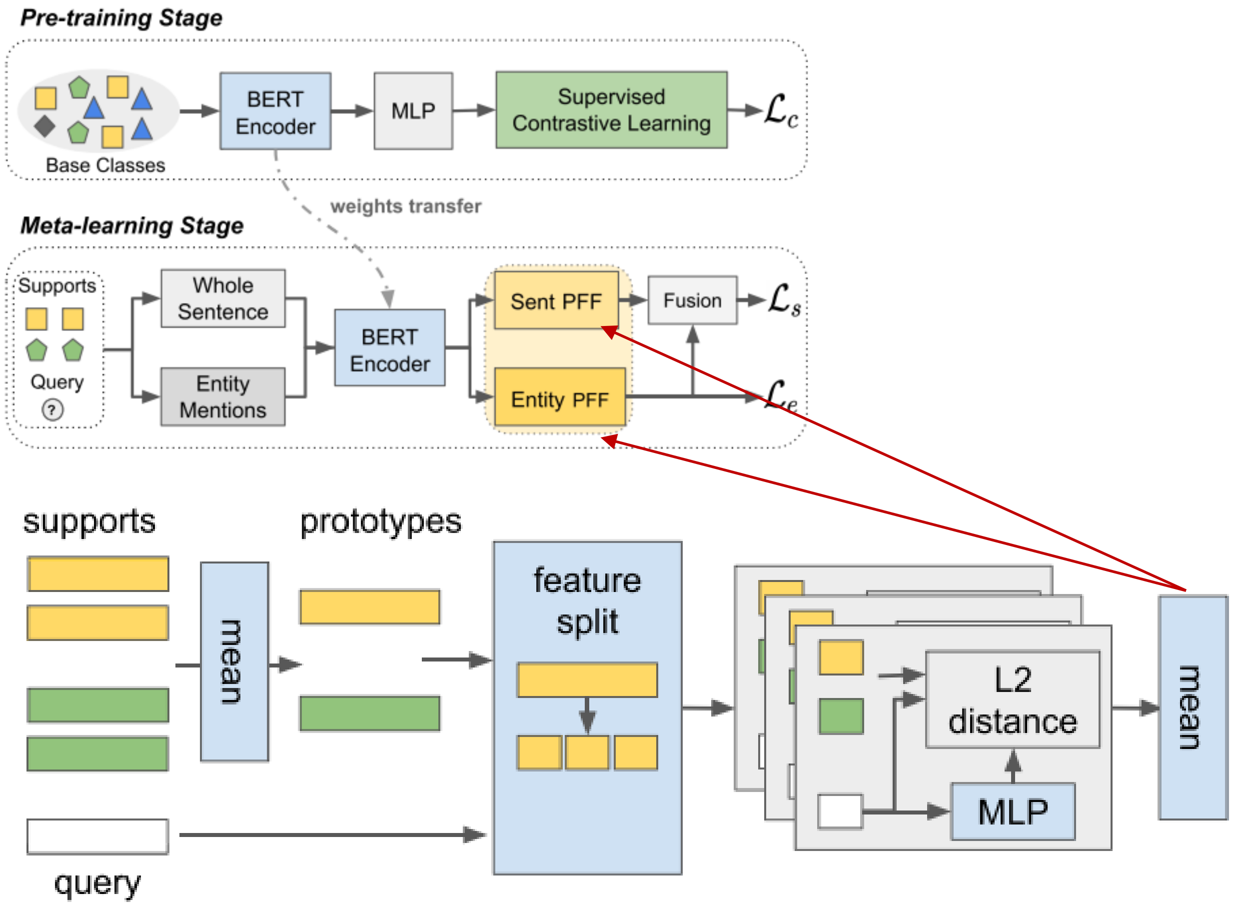
Pretraining stage contrastive loss

$$\mathcal{L}_c = \sum_{i=1}^{B} \mathcal{L}_c^i$$

$$\mathcal{L}_c^i = \frac{-1}{N_{y_i} - 1} \sum_{j=1}^{B} \mathbb{I}_{i \neq j} \cdot \mathbb{I}_{y_i = y_j} \cdot \log \frac{\exp(v_i \cdot v_j / \tau)}{\sum_{k=1}^{B} \mathbb{I}_{i \neq k} \cdot \exp(v_i \cdot v_k / \tau)}$$

a mini-batch with B pieces of data

Sentence- and Entity-level Prototype-based Networks



Prototypical networks with fine-grained featurewise fusion (PFF)

- average feature of K support instances, average feature of N query instances,
- cut the feature dimension into C parts
- attention score vector by MLP for every part L2 distance
- the final probability of query q with class i
- L = Ls + λ × Le

Experiments

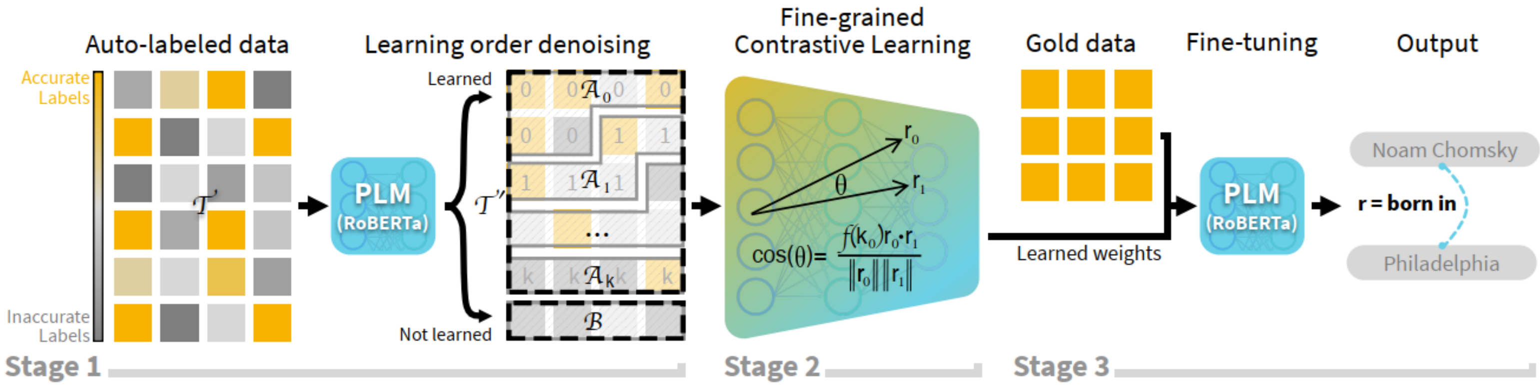
Model	5-way-1-shot	5-way-5-shot	10-way-1-shot	10-way-5-shot
Proto-CNN [30]	72.65 / 74.52	86.15 / 88.40	60.13 / 62.38	76.20 / 80.45
Proto-BERT [30]	82.92 / 80.68	91.32 / 89.60	73.24 / 71.48	83.68 / 82.89
Proto-HATT [11]	75.01 / — —	87.09 / 90.12	62.48 / — —	77.50 / 83.05
MLMAN [43]	79.01 / 82.98	88.86 / 92.66	67.37 / 75.59	80.07 / 87.29
BERT-PAIR [13]	85.66 / 88.32	89.48 / 93.22	76.84 / 80.63	81.76 / 87.02
REGRAB [29]	**87.95** / 90.30	92.54 / 94.25	80.26 / 84.09	86.72 / 89.93
CTEG [39]	84.72 / 88.11	92.52 / 95.25	76.01 / 81.29	84.89 / 91.33
Ours	87.21 / **90.40**	**94.86** / **96.95**	**80.34** / **84.68**	**91.36** / **94.15**
Human	— — / 92.22	–	— — / 85.88	–

Accuracy(%) of few-shot classification on the FewRel 1.0 validation / test set

Model	5-way-1-shot	5-way-5-shot	10-way-1-shot	10-way-5-shot
Proto-CNN [30]	35.09 (↓ 39.43)	49.37 (↓ 39.03)	22.98 (↓ 39.40)	35.22 (↓ 45.23)
Proto-BERT [30]	40.12 (↓ 40.56)	51.50 (↓ 38.10)	26.45 (↓ 45.03)	36.93 (↓ 45.96)
Proto-ADV [13]	42.21	58.71	28.91	44.35
BERT-PAIR [13]	67.41 (↓ 20.91)	78.57 (↓ 14.65)	54.89 (↓ 25.74)	66.85 (↓ 20.17)
Ours	**79.33** (↓ 11.07)	**91.59** (↓ 5.36)	**67.48** (↓ 17.20)	**85.70** (↓ 8.45)

Accuracy(%) of few-shot classification on the FewRel 2.0 domain adaptation test set

Learning order denoising

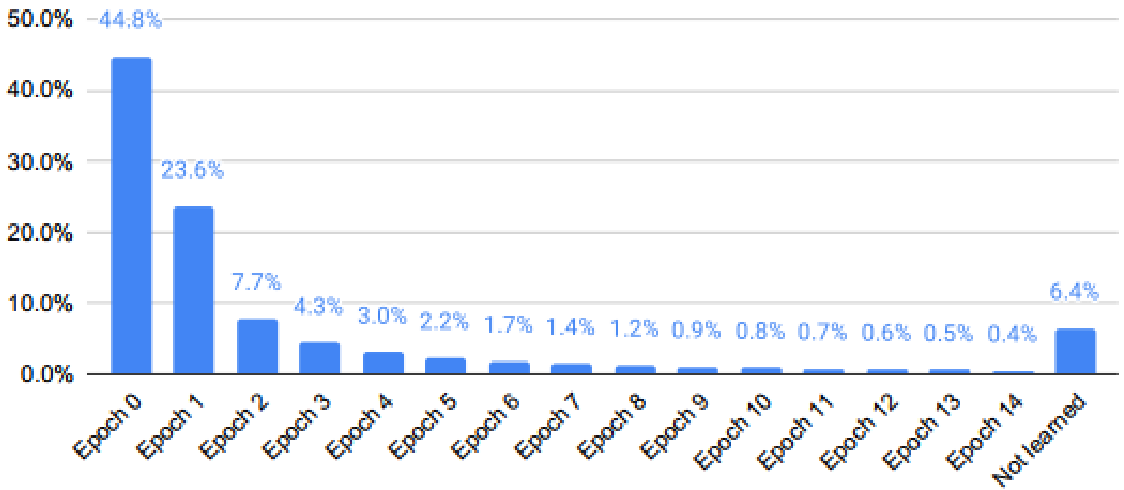


Weighted contrastive loss

$$\mathcal{L}_{\mathrm{RD}} = -\sum_{t_A, t_B \in \mathcal{T}'} f(k_A) \log \frac{\exp\left(\cos\left(\mathbf{r}_{t_A}, \mathbf{r}_{t_B}\right)/\tau\right)}{\mathcal{Z}}$$

$$\mathcal{Z} = \sum_{t_C \in \mathcal{T}'/\{t_A\}}^{N} f(k_C) \exp\left(\cos\left(\mathbf{r}_{t_A}, \mathbf{r}_{t_C}\right)/\tau\right),$$

$$f(k) = \alpha^{\frac{k_{\max} - k}{k_{\max} - k_{\min}}}$$



Percent of total training instances learned per epoch

Learning order	Training set	Training set size	F1	IgF1
None	$\mathcal{T}$	100%	**56.4**	45.8
Batch-based	$\mathcal{T}'_{\mathcal{A}_0^B}$	45.0%	**56.4**	**46.6**
Epoch-based	$\mathcal{T}'_{\mathcal{A}_0^E}$	64.9%	**56.4**	46.0

Results comparing performance on the DocRED test set by the instances learned in the first epoch.

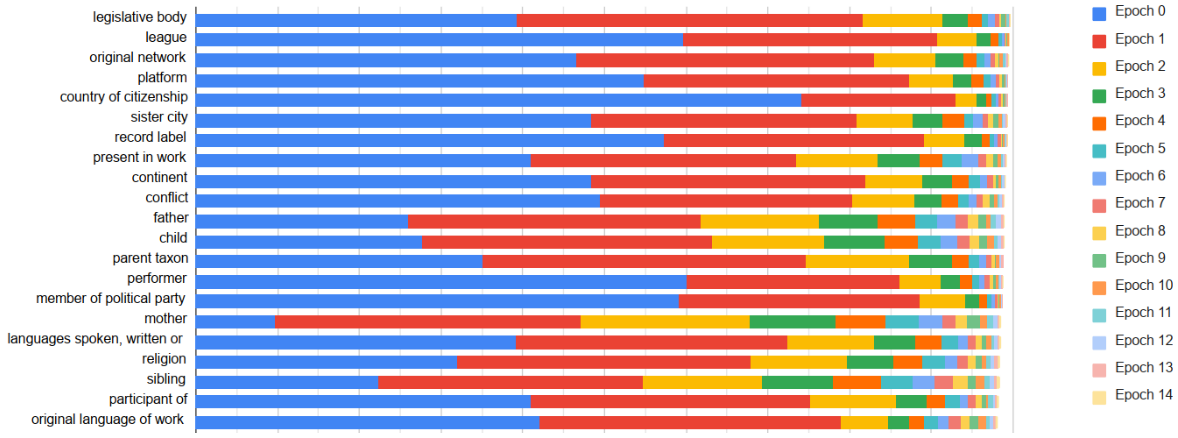
https://arxiv.org/abs/2205.12491

Experiments

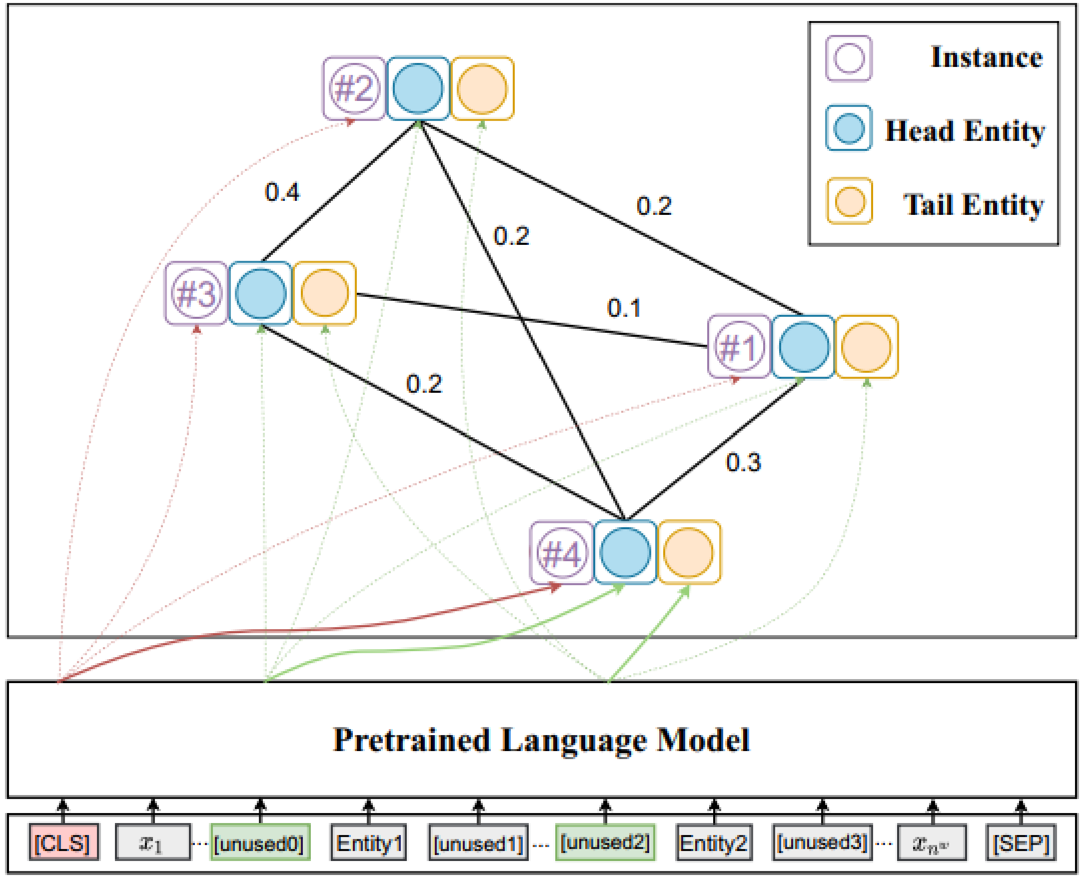
Size	1%		10%		100%	
Metrics	F1	IgF1	F1	IgF1	F1	IgF1
CNN*	-	-	-	-	42.3	40.3
BiLSTM*	-	-	-	-	51.1	50.3
HINBERT*	-	-	-	-	55.6	53.7
CorefBERT*	32.8	31.2	46.0	43.7	57.0	54.5
SpanBERT*	32.2	30.4	46.4	44.5	57.3	55.0
ERNIE*	26.7	25.5	46.7	44.2	56.6	54.2
MTB*	29.0	27.6	46.1	44.1	56.9	54.3
CP*	30.3	28.7	44.8	42.6	55.2	52.7
BERT	19.9	18.8	45.2	43.1	56.6	54.4
RoBERTa	29.6	27.9	47.6	45.7	58.2	55.9
ERICA$_{BERT}$	22.9	21.7	48.5	46.4	57.4	55.2
ERICA$_{RoBERTa}$	30.0	28.2	50.1	48.1	59.1	56.9
FineCL	**33.2**	**31.2**	**50.3**	**48.3**	**59.4**	**57.1**

IgF1 ignores performance on fact triples in the test set overlapping with triples in the train/dev sets DocRED

Dataset	TACRED			SemEval		
Size	1%	10%	100%	1%	10%	100%
MTB*	35.7	58.8	68.2	44.2	79.2	88.2
CP*	37.1	60.6	68.1	40.3	80.0	88.5
BERT	22.2	53.5	63.7	41.0	76.5	87.8
RoBERTa	27.3	61.1	69.3	43.6	77.7	87.5
ERICA$_{BERT}$	34.9	56.0	64.9	46.4	79.8	88.1
ERICA$_{RoBERTa}$	41.1	61.7	69.5	50.3	80.9	88.4
FineCL	**43.7**	**62.7**	**70.3**	**51.2**	**81.0**	**88.7**



BERT-based Graph convolutional network Model



instance bag *B*(*eh*,*et*)

{[CLS], x_1, x_2, [*unused*0], e_h, [*unused*1], · · ·, [*unused*2], e_t, [*unused*3], x_{nw}, [*SEP*]*}*

$$\mathcal{A} = \operatorname{softmax}\left(\frac{QW^Q \times \left(KW^K\right)^{\mathrm{T}}}{\sqrt{d_h}}\right)$$

where *Q* and *K* are both the concatenated representations of instances

$$\hat{r} = \operatorname{softmax}\left(\operatorname{MLP}(\operatorname{AvgPooling}(H^{(L)}))\right)$$

Experiments

Method	P@N							AUC
	100	200	300	500	1000	2000	MEAN	
PCNN+ATT (Lin et al., 2016)	73.0	68.0	67.3	63.6	53.3	40.0	60.9	34.1
BGWA (Jat et al., 2017)	76.0	74.0	-	-	-	-	-	36.7
CNN+RL (Feng et al., 2018)	79.0	73.0	-	-	-	-	-	37.4
DSGAN (Qin et al., 2018)	80.0	78.0	-	-	-	-	-	38.0
RESIDE (Vashishth et al., 2018)	81.8	75.4	74.3	69.7	59.3	45.0	67.6	41.5
PCNN+HATT (Han et al., 2018)	82.0	79.5	75.3	67.0	57.7	41.9	67.2	42.0
PCNN+BAG_ATT (Ye and Ling, 2019)	**91.8**	83.0	76.3	70.2	52.0	34.2	67.9	42.2
PA-TMR (Kuang et al., 2020)	83.0	79.0	-	-	-	-	-	43.7
ToHRE (Yu et al., 2020)	91.5	82.9	79.6	**74.8**	63.3	48.9	73.5	-
PA-TRP (Cao et al., 2021)	87.0	79.5	77.3	68.6	59.0	44.6	67.9	41.5
SRKBP (Christopoulou et al., 2021)	83.0	75.5	73.0	-	-	-	-	42.9
PSAN-RE (Shang et al., 2022)	79.2	71.1	66.8	65.9	60.4	48.1	65.2	43.8
DISTRE (Alt et al., 2019) ♯	68.0	67.0	65.3	65.0	60.2	47.9	62.2	42.2
REDSandT (Christou and Tsoumakas, 2021) ♯	78.0	-	73.0	67.6	-	-	-	42.9
CIL (Chen et al., 2021) ♯	90.1	86.1	**81.8**	-	-	-	-	50.8
Our BGM	90.3	**86.5**	80.0	74.6	**67.5**	**50.7**	**74.9**	**51.5**

Comparison results on NYT10

Method	P@100	P@200	AUC
PCNN+ATT (Lin et al., 2016)	94.0	93.0	80.3
BGWA (Jat et al., 2017)	99.0	98.0	81.5
CNN+RL (Feng et al., 2018)	**100.0**	96.0	85.5
DSGAN (Qin et al., 2018)	99.0	97.0	84.5
RESIDE (Vashishth et al., 2018)	**100.0**	97.5	89.1
PCNN+HATT (Han et al., 2018)	99.0	97.0	85.4
PA-TMR (Kuang et al., 2020)	**100.0**	98.0	86.5
PA-TRP (Cao et al., 2021)	**100.0**	98.0	87.3
PSAN-RE (Shang et al., 2022)	97.0	**98.5**	**91.1**
Our BGM	**100.0**	98.0	89.2

Comparison results on GDS

PLM	AUC	P@M	F1
Bert-based-uncased	51.5	**74.9**	52.4
Bert-based-cased	49.7	71.0	53.7
Bert-large-uncased	52.9	72.4	56.3
Distilbert-base-uncased	49.5	71.5	50.9
Xlnet-base-cased	47.5	67.1	53.1
Albert-based-v2	48.3	72.1	50.3
Roberta-base	48.8	68.5	52.7
Roberta-large	**53.2**	74.5	**57.1**

Comparison of PLMs in BGM on NYT10

Few-shot + weakly

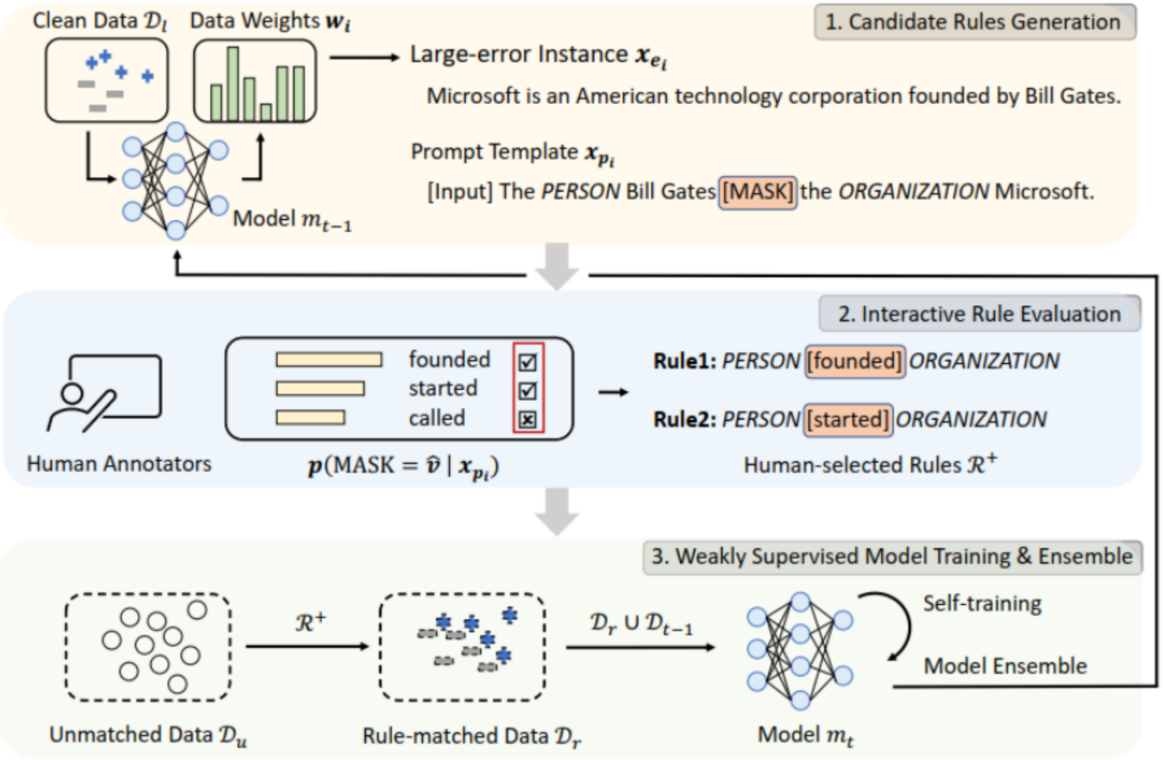


Figure 1: Overall framework for PRBOOST. In each iteration, PRBOOST (1) identifies large-error instances from the limited clean data and converts each large-error instance to a prompt template for prompting-based rule discovery; (2) presents candidate rules to human experts for annotation and uses accepted rules to generate new weak labels; (3) trains a new weak model with self-training and ensembles it with the previous models.

- Candidate rules generation

Input :	Microsoft is an American technology corporation founded by Bill Gates.
Prompt :	`[Input]` The Person Bill Gates `[Mask]` the Organization Microsoft.
Rule :	{Entity Pair == (Person, Org)} $\wedge$ { `[Mask]` == *founded*} $\wedge$ {$s_{t,j} \geq$ threshold} $\rightarrow$ **per:found**
Input :	Marvell Software Solutions Israel is a wholly owned subsidiary of Marvell Technology Group.
Prompt :	`[Input]` The Marvell Software Solutions Israel is a `[Mask]`.
Rule :	{ `[Mask]` == *subsidiary* $\vee$ *corporation* $\vee$ *company*} $\wedge$ {$s_{t,j} \geq$ threshold} $\rightarrow$ **Company**
Input :	Liverpool short of firepower for crucial encounter. Rafael Benitez must gamble with Liverpools Champions League prospects tonight but lacks the ammunition to make it a fair fight.
Prompt :	`[Mask]` News: `[Input]`
Rule :	{ `[Mask]` == *Liverpool* $\vee$ *Team* $\vee$ *Football* $\vee$ *Sports*} $\wedge$ {$s_{t,j} \geq$ threshold} $\rightarrow$ **Sports**

Table 1: The examples of prompt-based rules for relation extraction, ontology classification, and news topic classification. Here `[Input]` denotes the original input, `[Mask]` denotes the mask token, and $\wedge$, $\vee$ are the logical operators. We use bold words to show the ground-truth label of the original input.

The examples of prompt-based rules for relation extraction, ontology classification, and news topic classification

$$p(\mathrm{MASK} = \hat{\mathbf{v}} \mid \boldsymbol{x}_{p_i}) = \frac{\exp\left(\hat{\mathbf{v}} \cdot \mathcal{M}(\boldsymbol{x}_{p_i})\right)}{\sum_{\mathbf{v} \in \mathcal{V}} \exp\left(\mathbf{v} \cdot \mathcal{M}(\boldsymbol{x}_{p_i})\right)}$$

- Interactive rule evaluation

Few-shot + weakly

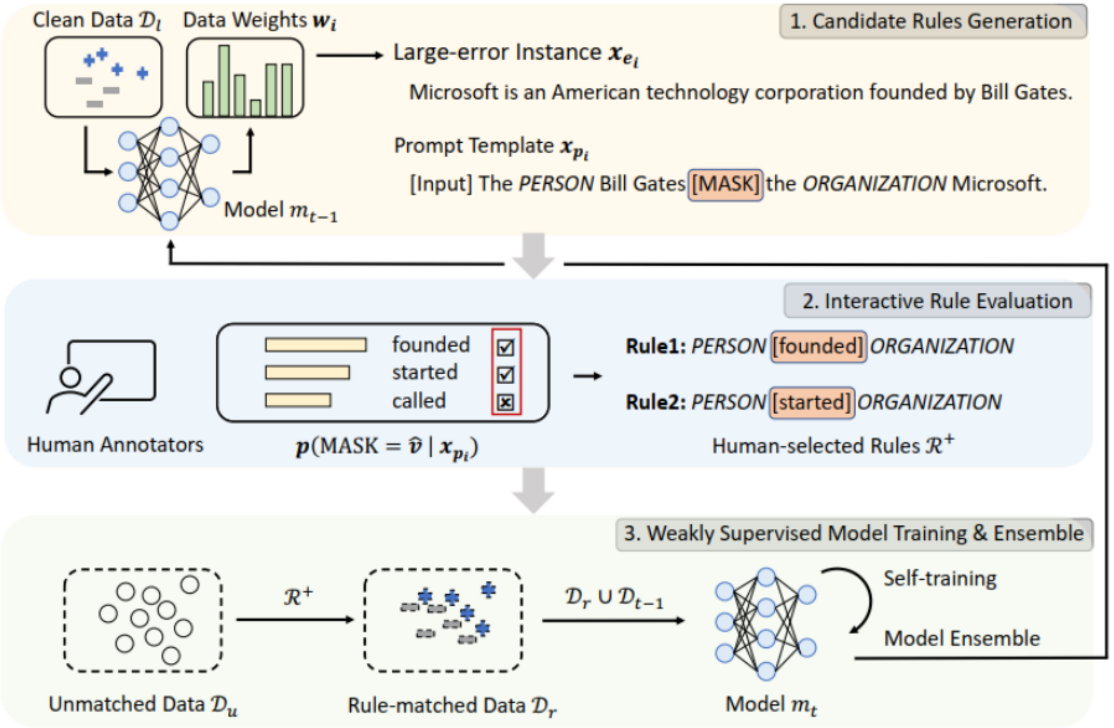


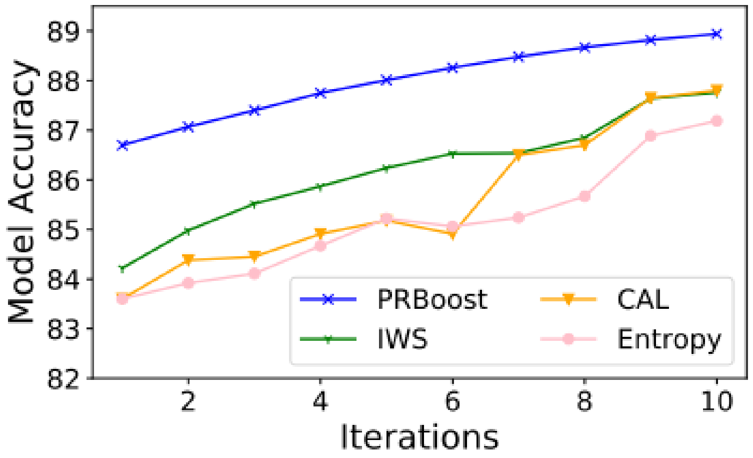
Figure 1: Overall framework for PRBOOST. In each iteration, PRBOOST (1) identifies large-error instances from the limited clean data and converts each large-error instance to a prompt template for prompting-based rule discovery; (2) presents candidate rules to human experts for annotation and uses accepted rules to generate new weak labels; (3) trains a new weak model with self-training and ensembles it with the previous models.

- Weaky supervised model training
  - embedding-based similarity
  - prompt-based vocabulary similarity
  - matching score
  - weakly labeled dataset cross entropy loss
  - self-training technique

PRBOOST: Prompt-Based Rule Discovery and Boosting for Interactive Weakly-Supervised Learning (ACL 2022)

Experiments

Method (Metrics)	TACRED (F1)	DBpedia (Acc.)	ChemProt (Acc.)	AG News (Acc.)
Supervised Baselines				
PLM w. 100% training data	66.9 (66.3/67.6)	99.4	79.7	94.4
PLM w. limited training data†	32.9 (40.8/27.6)	98.0	59.4	86.4
Weakly Supervised Baselines				
Rule Matching	20.1 (**85.0**/11.4)	63.2	46.9	52.3
Snorkel (Ratner et al., 2017)	39.7 (39.2/40.1)	69.5	56.4	86.2
LOTClass (Meng et al., 2020)	—	91.1	—	86.4
COSINE (Yu et al., 2021b)	39.5 (38.9/40.3)	73.1	59.8	87.5
Snorkel + fine-tuning†	40.8 (41.0/40.6)	97.6	64.9	87.7
LOTClass + fine-tuning†	—	98.1	—	88.0
COSINE + fine-tuning†	41.0 (40.4/41.7)	97.9	65.7	88.0
PRBoost	**48.1** (42.7/**55.1**)	**98.3**	**67.1**	**88.9**



Results of interactive methods on AG News

Continual Learning

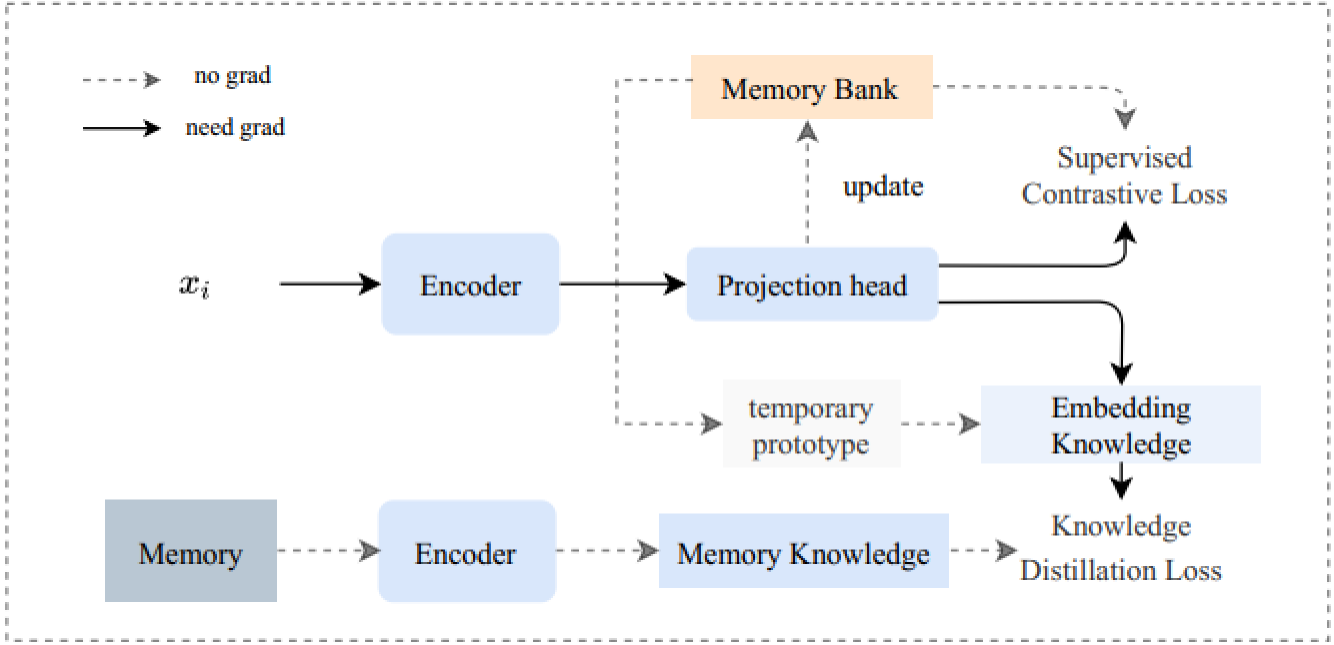


Figure 1: Framwork of consistent representation learning.

- Inital training for new task

$$\mathcal{L}_{\mathrm{CL}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(z_i \cdot z_p / \tau\right)}{\sum_{j \in S_I} \exp\left(z_i \cdot z_j / \tau\right)},$$

After backpropagating the gradient of loss

$$M_b[\tilde{I}] \leftarrow \{\mathbf{z}_i\}_{i=1}^{|B|}.$$

Continual Learning

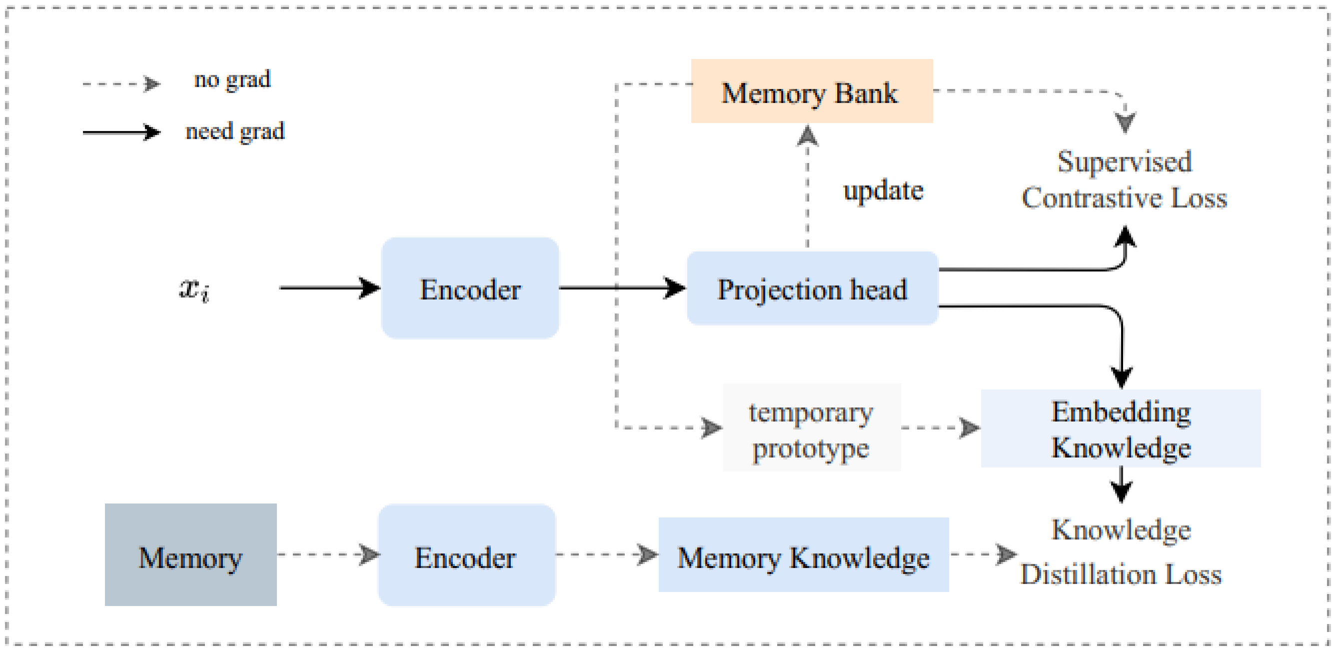


Figure 1: Framwork of consistent representation learning.

- Selecting Typical Samples for Memory
  - use k-means to cluster each relation
- Contrastive Replay with Memory Bank

$$\mathcal{L}_{\mathrm{CR}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(z_i \cdot z_p / \tau\right)}{\sum_{j \in \tilde{S}_I} \exp\left(z_i \cdot z_j / \tau\right)}$$

- Knowledge Distillation for Relieve Forgetting
  - the prototype before this task training
  - the prototype after cl loss update model
  - two prototype KL loss

- Prediction

$$p_c = \frac{1}{n_c} \sum_i \mathbf{E}\left(\bar{x}_i\right) \cdot \mathbb{1}\left\{y_i = c\right\},$$

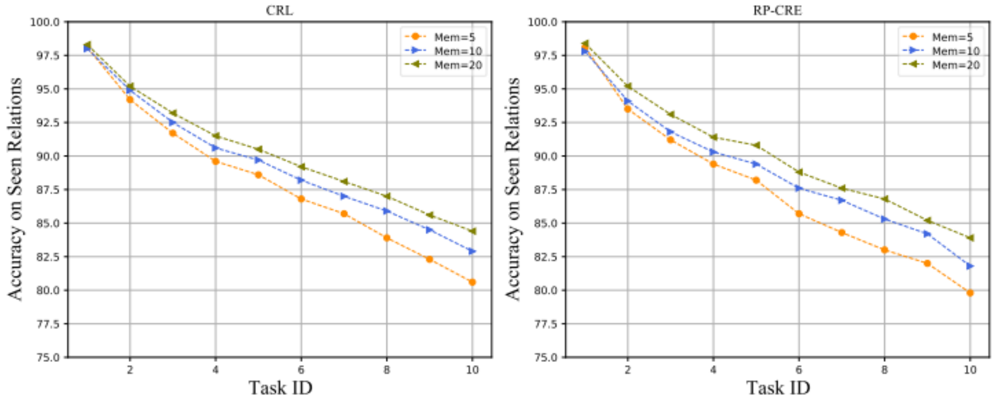
$$y^* = \underset{c=1,\ldots,k}{\operatorname{argmin}} \left\| f(\mathrm{x}) - p_c \right\|,$$

$$a_{ij} = \frac{p_i^T p_j}{\|p_i\| \, \|p_j\|},$$

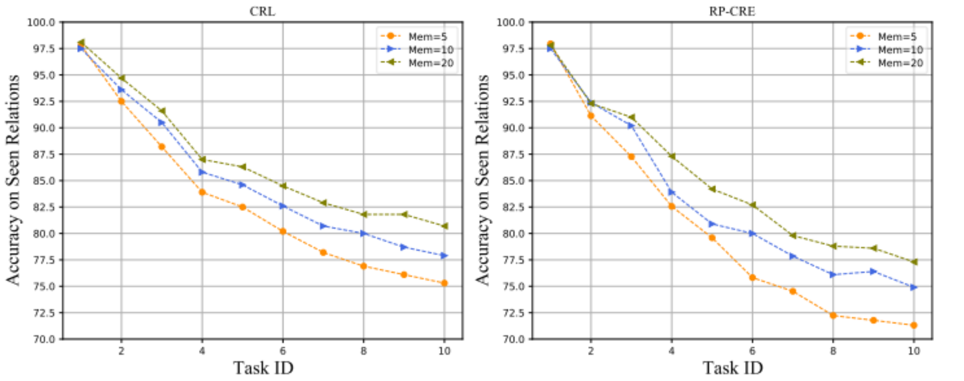
Experiments

FewRel										
Model	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EA-EMR	89.0	69.0	59.1	54.2	47.8	46.1	43.1	40.7	38.6	35.2
EMAR	88.5	73.2	66.6	63.8	55.8	54.3	52.9	50.9	48.8	46.3
CML	91.2	74.8	68.2	58.2	53.7	50.4	47.8	44.4	43.1	39.7
EMAR+BERT	**98.8**	89.1	89.5	85.7	83.6	84.8	79.3	80.0	77.1	73.8
RP-CRE	97.9	92.7	91.6	89.2	88.4	86.8	85.1	84.1	82.2	81.5
RP-CRE†	97.8	**95.1**	91.8	90.5	**89.9**	87.7	86.6	85.6	84.3	82.6
CRL	98.2	94.6	**92.5**	**90.5**	89.4	**87.9**	86.9	**85.6**	**84.5**	**83.1**
w/o KL	98.2	94.6	92.4	90.5	89.5	87.7	**87.1**	85.4	84.2	82.7
w/o CR	98.2	94.7	92.0	90.2	88.9	87.1	85.8	84.6	83.0	81.5

TACRED										
Model	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EA-EMR	47.5	40.1	38.3	29.9	24	27.3	26.9	25.8	22.9	19.8
EMAR	73.6	57.0	48.3	42.3	37.7	34.0	32.6	30.0	27.6	25.1
CML	57.2	51.4	41.3	39.3	35.9	28.9	27.3	26.9	24.8	23.4
EMAR+BERT	96.6	85.7	81	78.6	73.9	72.3	71.7	72.2	72.6	71.0
RP-CRE	97.6	90.6	86.1	82.4	79.8	77.2	75.1	73.7	72.4	72.4
RP-CRE†	97.6	93.1	**90.6**	**85.1**	82.7	81.1	78.3	76.0	76.1	75.7
CRL	**97.7**	93.2	89.8	84.7	84.1	81.3	**80.2**	79.1	**79.0**	**78.0**
w/o KL	97.7	**94.3**	90.1	84.9	**84.7**	**82.5**	80.0	**79.2**	79.0	77.7
w/o CR	97.7	92.7	88.8	84.7	82.3	80.5	77.8	75.9	75.2	74.3



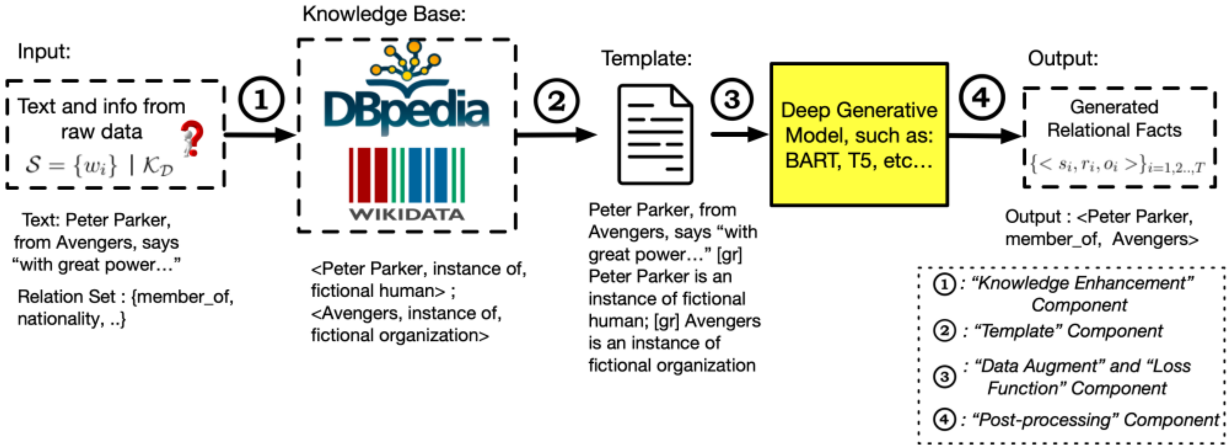
(a) Results on FewRel.



(b) Results on TACRED.

Unified view for relation extraction

Task name	**Entity Type aware Relation Classification**
Text †	Born in London in 1939 the son of a Greek tycoon, Negroponte grew up in Britain, Switzerland and the United States
Entity Positions	Negroponte: (51, 61) United States: (103, 116)
Entity Types	Negroponte: PERSON United States: LOCATION
Task name	**Relation Classification**
Text ‡	The cutting machine contains 13 circular blades mounted on a cutting axis.
Entity Positions	machine: (12, 19) blades: (41, 47)
Task name	**Joint Entity and Relation Extraction**
Text §	It is Japan's second-biggest automaker, behind Toyota and ahead of Nissan.



In the Relation Classification task， bi-encoder entity linking method (BLINK) links entities to Wikipedia

$$\mathcal{T}_1(\mathcal{S}|\mathcal{K}_\mathcal{S}) = \{w_1, .., w_{i-1}, [es], w_i^{s_t}, .., w_j^{s_t}, [gr], \mathbf{t}_{s_t}, w_{j+1}, .., w_{k-1}, [es], w_k^{o_t}, .., w_l^{o_t}, [gr], \mathbf{t}_{o_t}, w_{l+1}, .., w_N\},$$

In the task of Joint Entity and Relation Extraction,

$$\mathcal{T}_2(\mathcal{S}|\mathcal{K}_\mathcal{S}) = \{w_1, w_2, .., w_N, [gr], \mathbf{e_1} is\ an\ instance\ of\ \mathbf{t}_{e_1}, ..., [gr], \mathbf{e}_M\ is\ an\ instance\ of\ \mathbf{t}_{e_M} ..\},$$

Experiments

generated in the text format in a single pass

$\{s_1\ r_1\ o_1;\ s_2\ r_2\ o_2;\ ...;\ s_T\ r_T\ o_T\}$

Post-processing

- generated entity with the gold entity Token ";" is used to separate triples in the training stage
- Relation Classification task ---Levenshtein similarity, replace the with the gold entity
- Joint Entity and Relation Classification ---match the generated entities to the set of subspans in the text

Dataset	Method	Prec.	Rec.	F1
TACRED	ERNIE (Zhang et al., 2019)	80.0	66.1	68.0
	SpanBERT (Joshi et al., 2020)	70.8	70.9	70.8
	K-Adapter (Wang et al., 2020a)	68.9	75.4	72.0
	RoBERTa (Wang et al., 2020a)	70.2	72.4	71.3
	LUKE (Yamada et al., 2020)	70.4	75.1	72.7
	RECENT (Lyu and Chen, 2021)	90.9	64.2	**75.2**
	REKnow	76.2 (0.51)	74.1 (0.45)	**75.1** (0.47)
Semeval	CR-CNN (Santos et al., 2015)	-	-	84.1
	BERT-Entity (Soares et al., 2019)	-	-	89.2
	BERT-MTB (Soares et al., 2019)	-	-	**89.5**
	REKnow	88.1 (0.30)	91.4 (0.59)	**89.8** (0.26)

Main result for entity position-aware case

Experiments

Dataset	Method	Precison	Recall	F1
NYT	NovelTagging (Zheng et al., 2017)	32.8	30.6	31.7
	MultiHead (Bekoulis et al., 2018)	60.7	58.6	59.6
	ETL-Sapn (Yu et al., 2019)	85.5	71.7	78.0
	Tplinker (Wang et al., 2020b)	91.4	92.6	92.0
	CopyRE* † (Zeng et al., 2018)	61	56.6	58.7
	CopyMTL* † (Zeng et al., 2020)	75.7	68.7	72.0
	TANL (Paolini et al., 2021) †	-	-	90.8
	REBEL (Huguet Cabot and Navigli, 2021) †	-	-	**93.4**
	CGT † (Ye et al., 2020)	94.7	84.2	89.1
	REKnow	93.1 (0.18)	94.1 (0.17)	**93.6** (0.17)
Webnlg	NovelTagging (Zheng et al., 2017)	52.5	19.3	28.3
	MultiHead (Bekoulis et al., 2018)	57.5	54.1	55.7
	ETL-Sapn (Yu et al., 2019)	84.3	82.0	83.1
	Tplinker (Wang et al., 2020b)	88.9	84.5	**86.7**
	CopyRE* † (Zeng et al., 2018)	37.7	36.4	37.1
	CopyMTL* † (Zeng et al., 2020)	58.0	54.9	56.4
	CGT † (Ye et al., 2020)	92.9	75.6	83.4
	REKnow	90.4 (0.17)	87.9 (0.33)	**89.1** (0.24)
ACE2005	Attention (Katiyar and Cardie, 2017)	-	-	55.9
	DYGIE (Luan et al., 2019)	-	-	63.2
	DYGIE++ (Wadden et al., 2019)	-	-	63.4
	Pure-Bb (Zhong and Chen, 2020)	-	-	66.7
	Pure-Alb (Zhong and Chen, 2020)	-	-	**69.0**
	REKnow	71.3 (0.81)	67.6 (0.30)	**69.4** (0.54)

Main result for entity position-absent case

Dataset	REKnow	w/o KG	w/o Text
NYT	93.6 (0.17)	93.2 (0.16)	89.1 (0.16)
WebNLG	89.1 (0.24)	88.3 (0.49)	67.8 (0.32)
ACE2005	69.4 (0.54)	69.2 (0.80)	9.2 (0.33)

Analysis of knowledge enhancement

Dataset	Found Ext. Info Ratio		
	Train	Val	Test
NYT10	1.75	1.76	1.76
WebNLG	0.93	0.91	0.97
ACE2005	0.75	0.71	0.66

“Found Ext. Info Ratio” represents the ratio between the size of entities in the text and the size of found external information in entity linking step

Thanks