Artificial Intelligence

EIDER: Empowering Document-level Relation Extraction with Efficient Evidence Extraction and Inference-stage Fusion

Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, Jiawei Han
University of Illinois at Urbana-Champaign, IL, USA
{xyiqing2, js2, shal2, yuningm2, hanj}@illinois.edu

Code： https://github.com/Veronicium/Eider

——ACL2022

2022. 04. 03

Reported by ChangJiang Hu

Leibniz-Institut für Sozialwissenschaften

Introduction

Head: Hero of the Day Tail: the United States Rel: [country of origin]
GT evidence sentences: [1,10] Extracted evidence: [1,10]

Original document as input: [1] Load is the sixth studio album by the American heavy metal band Metallica, released on June 4, 1996 by Elektra Records in the United States … [9] It was certified 5×platinum … for shipping five million copies in the United States. [10] Four singles—"Hero of the Day", "Until It Sleeps", "Mama Said", and "King Nothing" — were released as part of the marketing campaign for the album.
Prediction scores: NA: 17.63 country of origin: 14.79

Extracted evidence as input: [1] Load is the sixth studio album … released … in the United States … [10] Four singles — "Hero of the Day", … were released … for the album.
Prediction scores: country of origin: 18.31 NA: 13.45

Final prediction of our model: country of origin (✓)

Figure 1: A test sample in the DocRED dataset (Yao et al., 2019), where the i^{th} sentence in the document is marked with [i] at the start. Our model correctly predicts [1,10] as evidence, and if we only use the extracted evidence as input, the model can predict the relation "country of origin" correctly.

Method

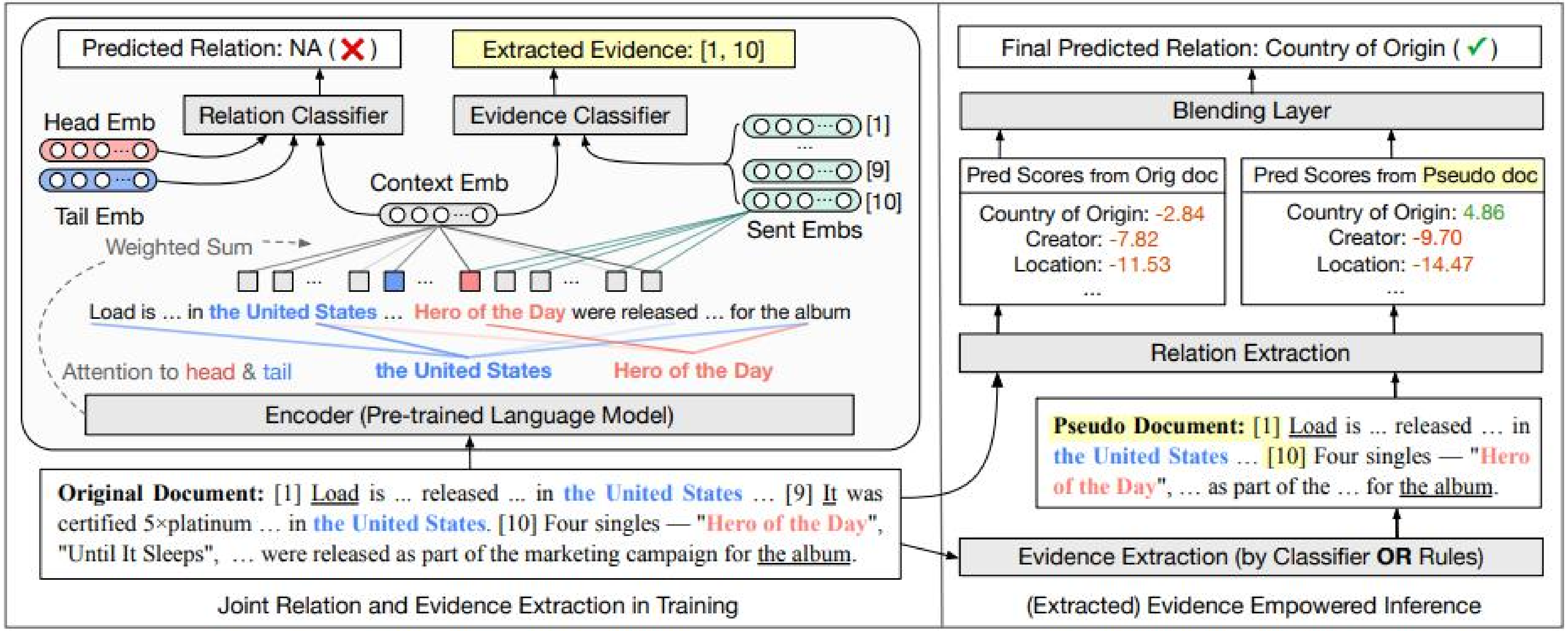


Figure 2: The overall architecture of EIDER. The left part illustrates the training stage and the right shows the inference stages of EIDER. We highlight head entities, tail entities and extracted evidences.

Method

Problem Formulation

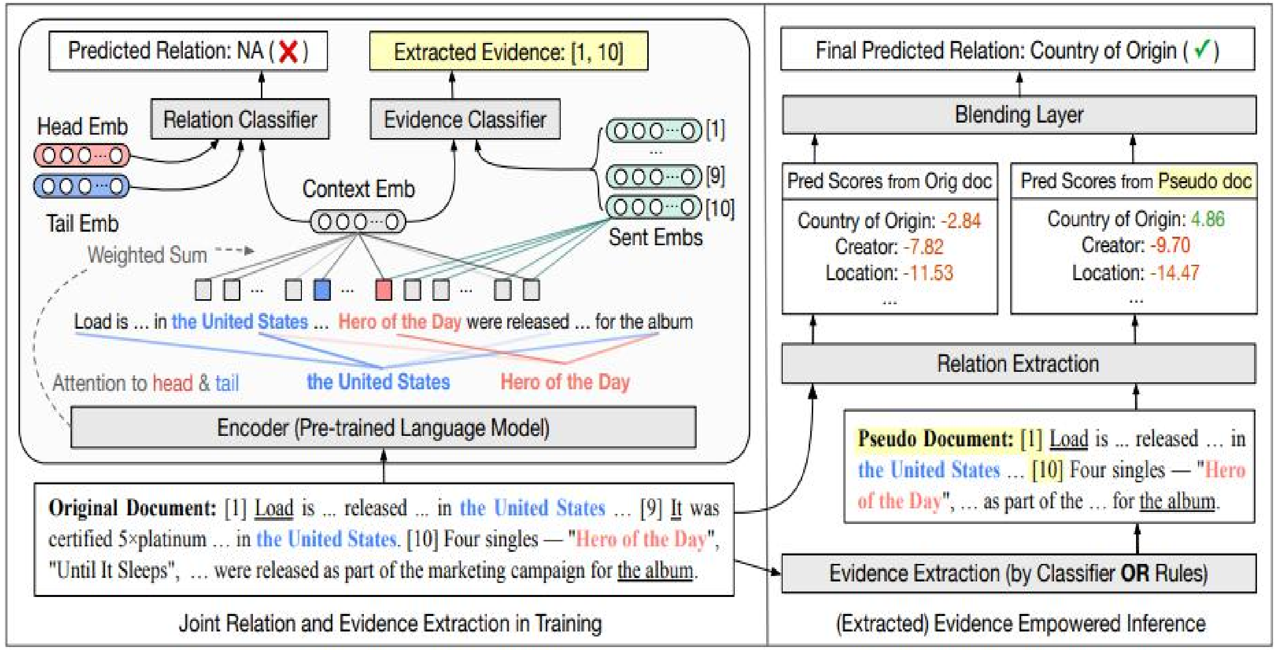


Figure 2: The overall architecture of EIDER. The left part illustrates the training stage and the right shows the inference stages of EIDER. We highlight head entities, tail entities and extracted evidences.

Document d

N sentences $\{s_n\}_{n=1}^{N}$

L tokens $\{h_l\}_{l=1}^{L}$

E named entities $\{e_i\}_{i=1}^{E}$ $\{m_j^i\}$

(e_h, e_t) $V_{h,t} = \{s_{v_k}\}_{k=1}^{K}$

Method

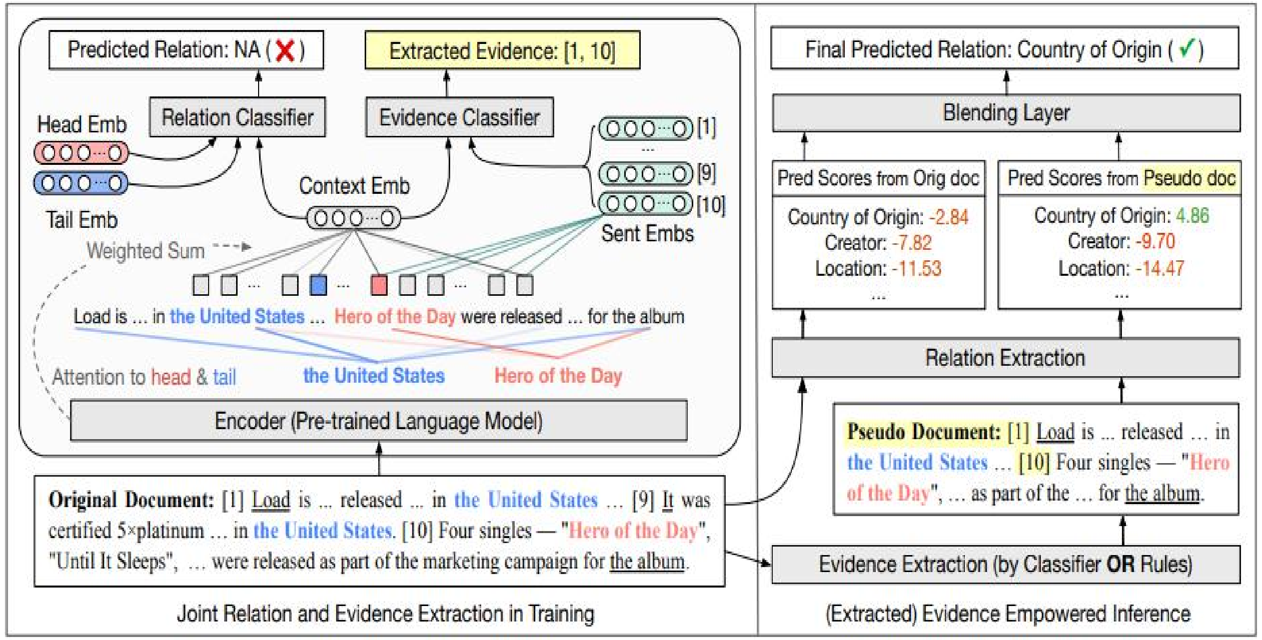


Figure 2: The overall architecture of EIDER. The left part illustrates the training stage and the right shows the inference stages of EIDER. We highlight head entities, tail entities and extracted evidences.

$$\boldsymbol{H}, \boldsymbol{A} = \mathrm{Encoder}([h_1, ..., h_L]), \qquad (1)$$

LogSumExp pooling $\mathbf{e}_i = \log \sum_j \exp(\mathbf{m}_\mathbf{j}^\mathbf{i}).$

$$\mathbf{c}_{h,t} = \boldsymbol{H}^T \frac{\boldsymbol{A}_h \circ \boldsymbol{A}_t}{\boldsymbol{A}_h^T \boldsymbol{A}_t}, \qquad (2)$$

Method

Relation Classifier

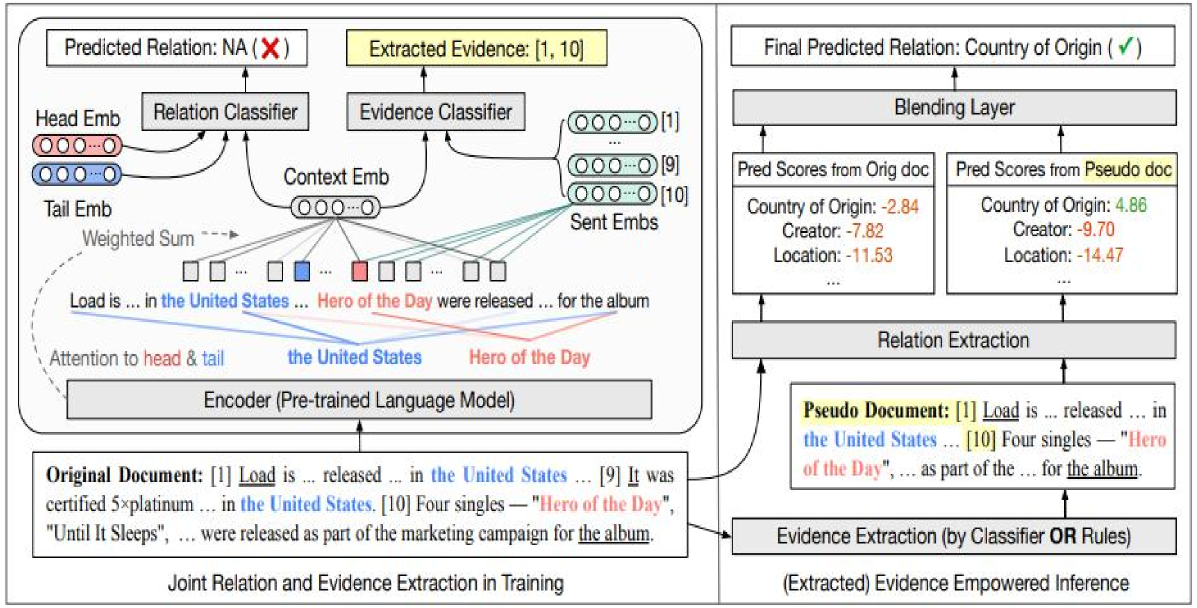


Figure 2: The overall architecture of EIDER. The left part illustrates the training stage and the right shows the inference stages of EIDER. We highlight head entities, tail entities and extracted evidences.

$$
\begin{aligned}
\boldsymbol{z}_h &= \tanh\left(\boldsymbol{W}_h \mathbf{e}_h + \boldsymbol{W}_{c_h} \mathbf{c}_{h,t}\right), \\
\boldsymbol{z}_t &= \tanh\left(\boldsymbol{W}_t \mathbf{e}_t + \boldsymbol{W}_{c_t} \mathbf{c}_{h,t}\right), \qquad (3)
\end{aligned}
$$

$$
\mathbf{y}_r = \boldsymbol{z}_h \boldsymbol{W}_r \boldsymbol{z}_t + \boldsymbol{b}_r,
$$

$$
\mathbf{y}_{\mathrm{TH}} = \boldsymbol{z}_h \boldsymbol{W}_{\mathrm{TH}} \boldsymbol{z}_t + \boldsymbol{b}_r. \qquad (4)
$$

$$
S^{(O)}_{h,t,r} = \mathbf{y}_r - \mathbf{y}_{TH}.
$$

$$
\begin{aligned}
\mathcal{L}_{RE} = -\sum_{h \neq t} \sum_{r \in \mathcal{P}^T_{h,t}} & \log\left(\frac{\exp\left(\mathbf{y}_r\right)}{\sum_{r' \in \mathcal{P}^T_{h,t} \cup \{\mathrm{TH}\}} \exp\left(\mathbf{y}_{r'}\right)}\right) \\
& - \log\left(\frac{\exp\left(\mathbf{y}_{\mathrm{TH}}\right)}{\sum_{r' \in \mathcal{N}^T_{h,t} \cup \{\mathrm{TH}\}} \exp\left(\mathbf{y}_{r'}\right)}\right). \qquad (5)
\end{aligned}
$$

Method

Evidence Classifier

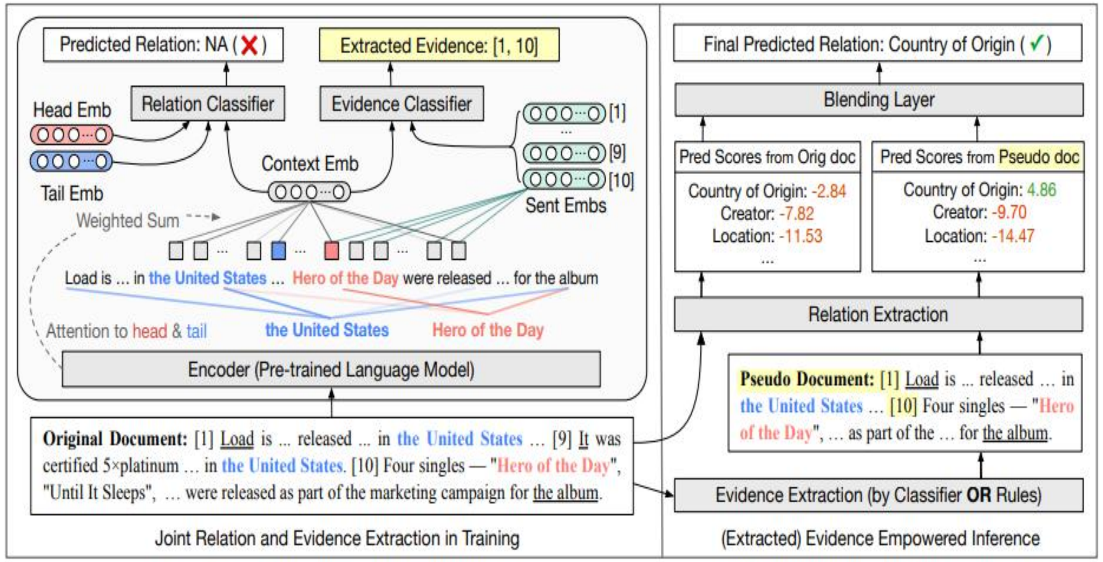


Figure 2: The overall architecture of EIDER. The left part illustrates the training stage and the right shows the inference stages of EIDER. We highlight head entities, tail entities and extracted evidences.

$$s_n\text{:}\ \mathbf{s}_n = \log \sum_{h_l \in s_n} \exp\left(\mathbf{h}_l\right).$$

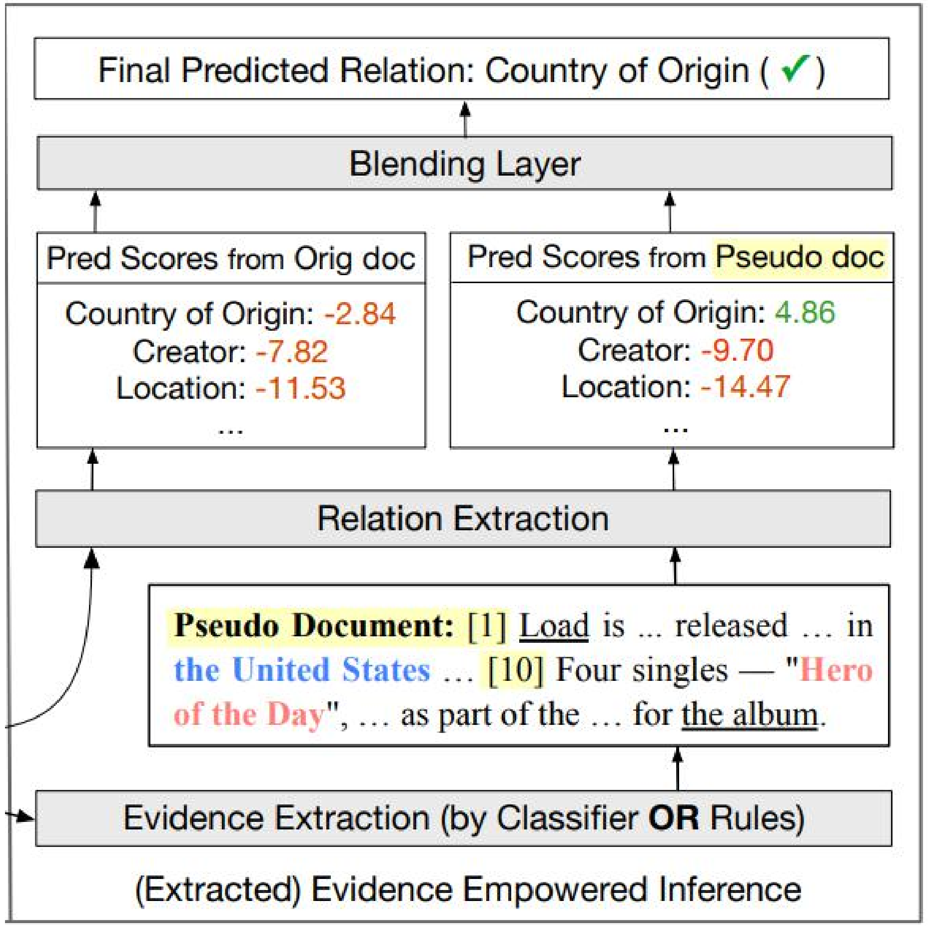
$$\mathrm{P}\left(s_n \mid e_h, e_t\right) = \sigma\left(\mathbf{s}_n \boldsymbol{W}_v \mathbf{c}_{h,t} + \boldsymbol{b}_v\right), \qquad (6)$$

$$\mathcal{L}_{Evi} = -\sum_{h \neq t, \mathrm{NA} \notin \mathcal{P}_{h,t}^{T}} \sum_{s_n \in \mathcal{D}} y_n \cdot \mathrm{P}\left(s_n \mid e_h, e_t\right) + \left(1-y_n\right) \cdot \log\left(1-\mathrm{P}\left(s_n \mid e_h, e_t\right)\right), \qquad (7)$$

$$\mathcal{L} = \mathcal{L}_{RE} + \mathcal{L}_{Evi}. \qquad (8)$$

Method

Fusion of Evidence in Inference



$$\mathrm{P}_{Fuse}(r|e_h, e_t) = \sigma(S^{(O)}_{h,t,r} + S^{(E)}_{h,t,r} - \tau). \quad (9)$$

$$\mathcal{L}_{Fuse} = -\sum_{d\in\mathcal{D}}\sum_{h\neq t}\sum_{r\in\mathcal{R}} y_r \cdot \mathrm{P}_{Fuse}(r|e_h, e_t) + (1 - y_r)\cdot\log(1 - \mathrm{P}_{Fuse}(r|e_h, e_t)), \quad (10)$$

Method

Heuristic Evidence Label Construction

Co-occur

Coref

Original Document: [1] Load is ... released ... in the United States … [9] It was certified 5×platinum … in the United States. [10] Four singles — "Hero of the Day", "Until It Sleeps", … were released as part of the marketing campaign for the album.

Bridge

Experiments

Model	Dev				Test	
	Ign F1	F1	Intra F1	Inter F1	Ign F1	F1
LSR-BERT$_{base}$ (Nan et al., 2020)	52.43	59.00	65.26	52.05	56.97	59.05
GLRE-BERT$_{base}$ (Wang et al., 2020)	-	-	-	-	55.40	57.40
Reconstruct-BERT$_{base}$ (Xu et al., 2021)	58.13	60.18	-	-	57.12	59.45
GAIN-BERT$_{base}$ (Zeng et al., 2020)	59.14	61.22	67.10	53.90	59.00	61.24
BERT$_{base}$ (Wang et al., 2019)	-	54.16	61.61	47.15	-	53.20
BERT-Two-Step (Wang et al., 2019)	-	54.42	61.80	47.28	-	53.92
HIN-BERT$_{base}$ (Tang et al., 2020)	54.29	56.31	-	-	53.70	55.60
E2GRE-BERT$_{base}$ (Huang et al., 2021a)	55.22	58.72	-	-	-	-
CorefBERT$_{base}$ (Ye et al., 2020)	55.32	57.51	-	-	54.54	56.96
ATLOP-BERT$_{base}$ (Zhou et al., 2021)	$59.11 \pm 0.14^{\dagger}$	$61.01 \pm 0.10^{\dagger}$	$67.26 \pm 0.15^{\dagger}$	$53.20 \pm 0.19^{\dagger}$	59.31	61.30
EIDER (Rule)-BERT$_{base}$	60.36 ± 0.13	62.34 ± 0.08	68.40 ± 0.14	54.79 ± 0.13	60.23	62.21
EIDER-BERT$_{base}$	**60.51 ± 0.11**	**62.48 ± 0.13**	**68.47 ± 0.08**	**55.21 ± 0.21**	**60.42**	**62.47**
RoBERTa$_{large}$ (Ye et al., 2020)	57.14	59.22	-	-	57.51	59.62
CorefRoBERTa$_{large}$ (Ye et al., 2020)	57.35	59.43	-	-	57.90	60.25
E2GRE-RoBERTa$_{large}$ (Huang et al., 2021a)	59.55	62.91	-	-	60.29	62.51
GAIN-BERT$_{large}$ (Zeng et al., 2020)	60.87	63.09	-	-	60.31	62.76
ATLOP-RoBERTa$_{large}$ (Zhou et al., 2021)	$61.30 \pm 0.22^{\dagger}$	$63.15 \pm 0.21^{\dagger}$	$69.61 \pm 0.25^{\dagger}$	$55.01 \pm 0.18^{\dagger}$	61.39	63.40
EIDER (Rule)-RoBERTa$_{large}$	61.73 ± 0.07	63.91 ± 0.07	69.99 ± 0.09	56.27 ± 0.11	61.93	64.12
EIDER-RoBERTa$_{large}$	**62.34 ± 0.14**	**64.27 ± 0.10**	**70.36 ± 0.07**	**56.53 ± 0.15**	**62.85**	**64.79**

Table 1: Relation extraction results on DocRED. We report the mean and standard deviation on the development set by conducting 5 runs with different random seeds. We report the official test score of the best checkpoint on the development set. Results with † are based on our implementation. Others are reported in their original papers. We separate graph-based and transformer-based methods into two groups.

Experiments

Model	CDR	GDA
LSR-BERT$_{base}$ (Nan et al., 2020)	64.8	82.2
SciBERT$_{base}$ (Zhou et al., 2021)	65.1 ± 0.6	82.5 ± 0.3
DHG-BERT$_{base}$ (Zhang et al., 2020b)	65.9	83.1
GLRE-SciBERT$_{base}$ (Wang et al., 2020)	68.5	-
ATLOP-SciBERT$_{base}$ (Zhou et al., 2021)	69.4 ± 1.1	83.9 ± 0.2
EIDER (Rule)-SciBERT$_{base}$	$\mathbf{70.63} \pm 0.49$	$\mathbf{84.54} \pm 0.22$

Table 2: Relation extraction results on CDR and GDA.

Experiments

Model	Dev Evi F1	Test Evi F1
E2GRE-BERT$_{base}$	47.14	48.35
EIDER-BERT$_{base}$	**50.71**	**51.27**
E2GRE-RoBERTa$_{large}$	51.11	50.50
EIDER-RoBERTa$_{large}$	**52.54**	**53.01**

Table 3: Evidence extraction results on DocRED. We compare EIDER with E2GRE (Huang et al., 2021a).

Experiments

	Rules (ours)	EIDER-BERT$_{base}$	NoJoint
PosEvi F1	77.43	**80.33**	51.13

Table 4: Ablation study for evidence extraction.

Ablation	Ign F1	F1	Intra F1	Inter F1
EIDER-BERT$_{base}$	**60.51**	**62.48**	**68.47**	**55.21**
NoJoint	59.98	62.03	68.51	54.10
NoPseudo	59.70	61.53	67.55	54.01
NoOrigDoc	58.47	60.44	66.24	53.23
NoBlending	58.93	61.46	67.33	54.37
FinetuneOnEvi	60.11	62.29	68.13	54.84
EIDER (Rule)-BERT$_{base}$	**60.36**	**62.34**	**68.40**	**54.79**
NoJoint	60.01	62.09	68.21	54.34

Table 5: Ablation study of EIDER on DocRED.

Experiments

	Co-occur	Coref	Bridge	Total
Count	6711	984	3212	10,907
Percent	54.46%	7.99%	26.07%	88.52%

Table 6: Statistics of the 12,323 relations in the DocRED development set.

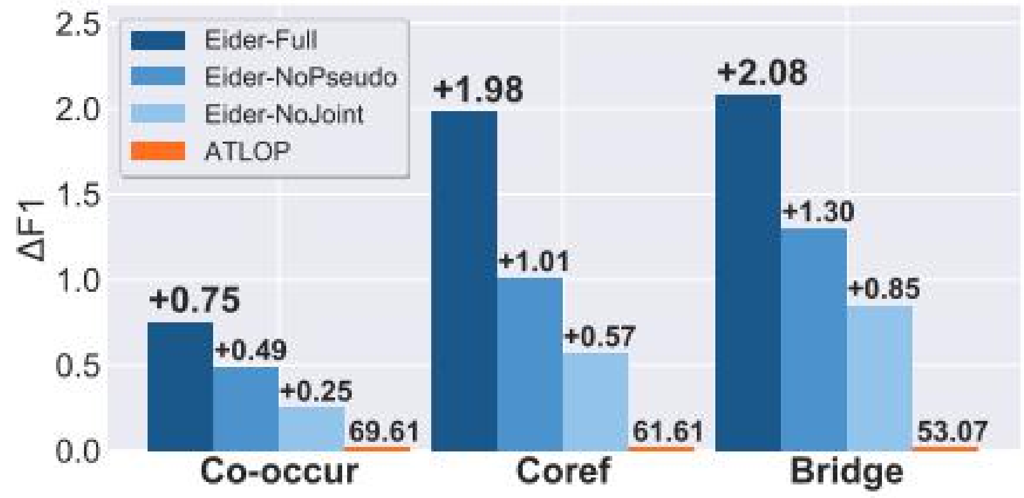


Figure 3: Performance gains in F1 by relation categories. The gains are relative to the second best baseline (ATLOP-RoBERTa$_{\text{large}}$).

Thank you!