Improving Named Entity Recognition by External Context Retrieving and Cooperative Learning

Xinyu Wang, Yong Jiang, Nguyen Bach, Tao Wang, Zhongqiang Huang, Fei Huang, Kewei Tu

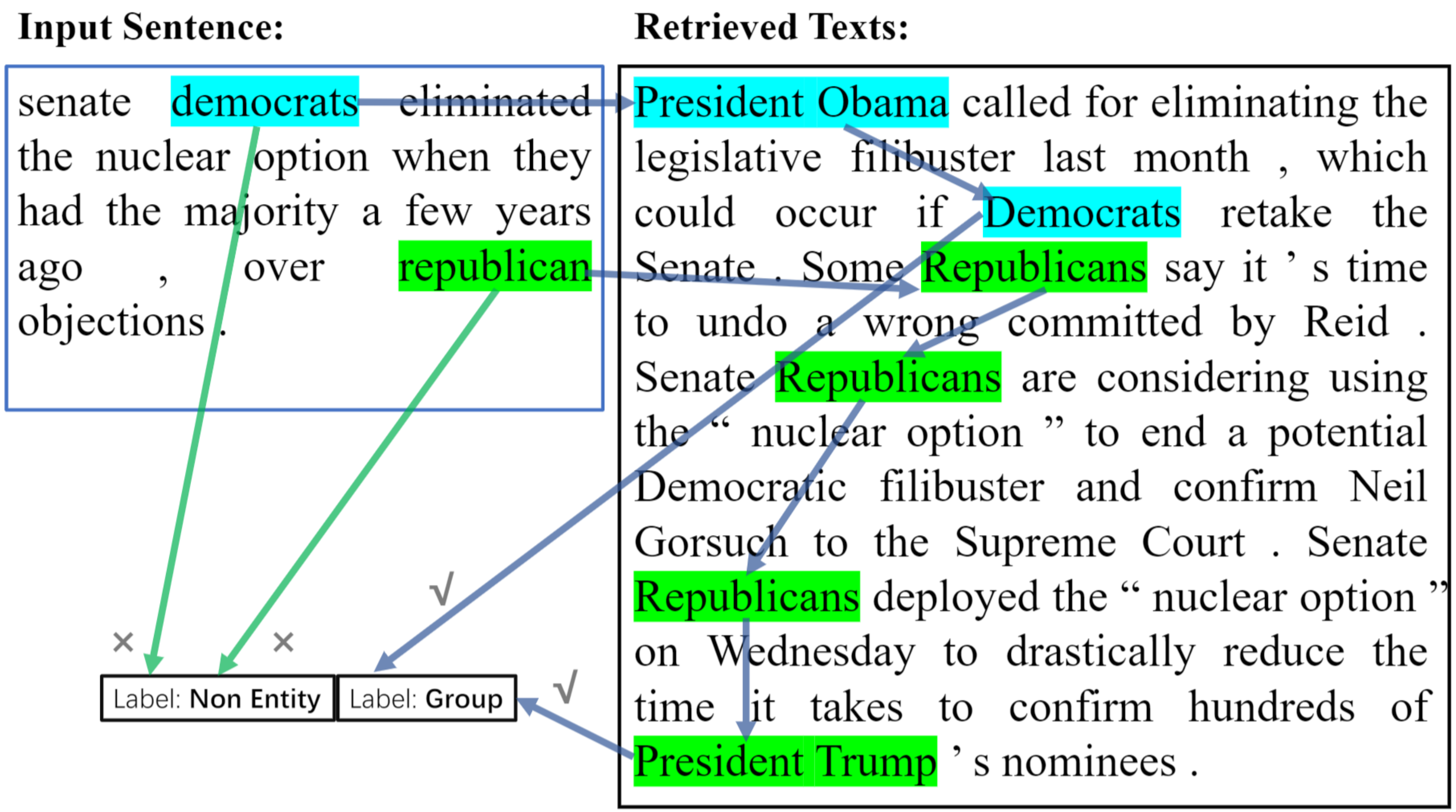
上海科技大学
ShanghaiTech University

DAMO
ALIBABA DAMO ACADEMY

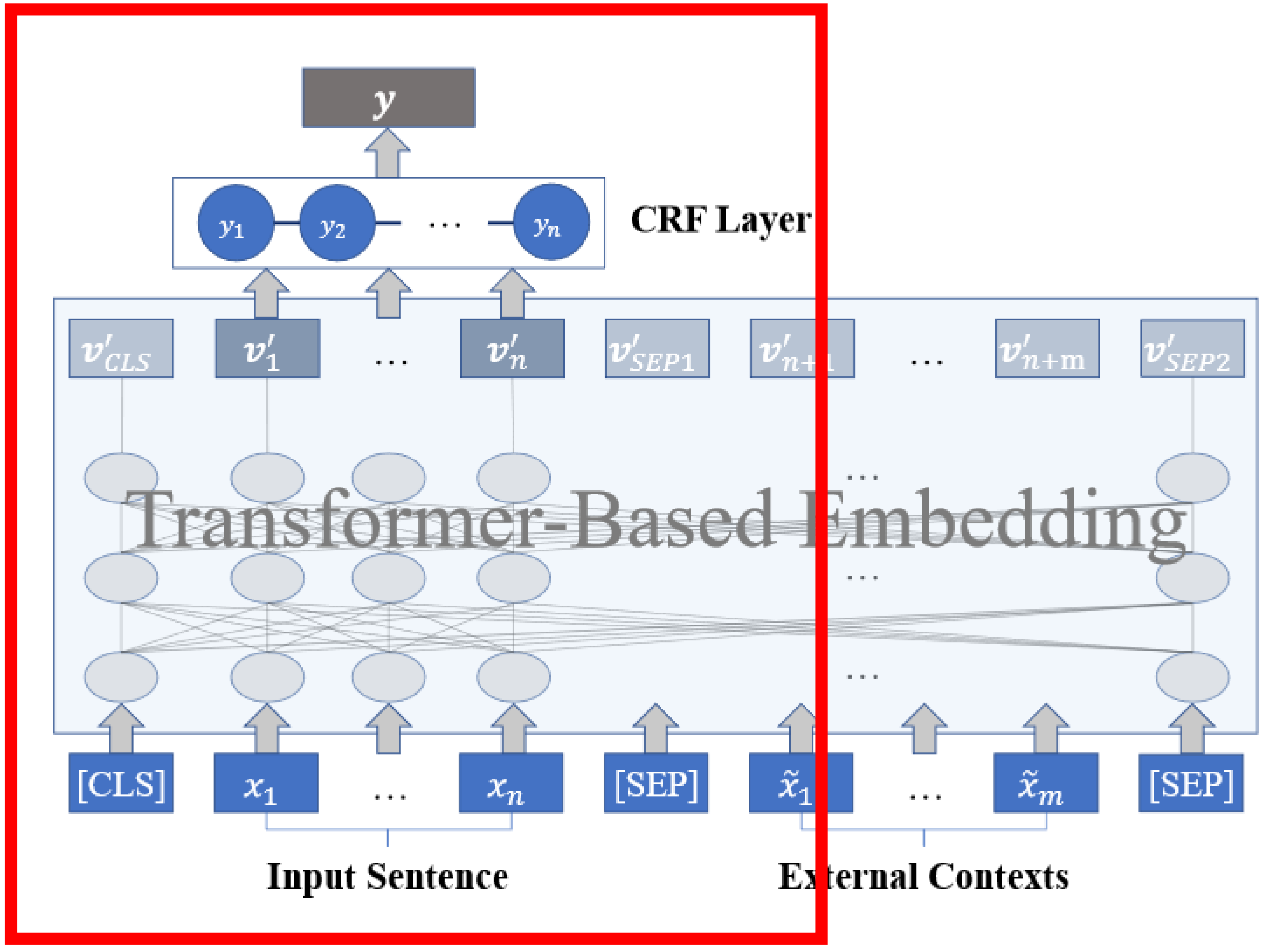
Motivation

- Previous studies
  - Including document-level contexts of the target sentence in the input of contextual embeddings methods can boost the accuracy of NER models
- Problem
  - There are a lot of application scenarios in which document-level contexts are unavailable
- Idea
  - Knowledge from search engines helps human annotators of NER
  - It should benefit machines too!

Motivating Example

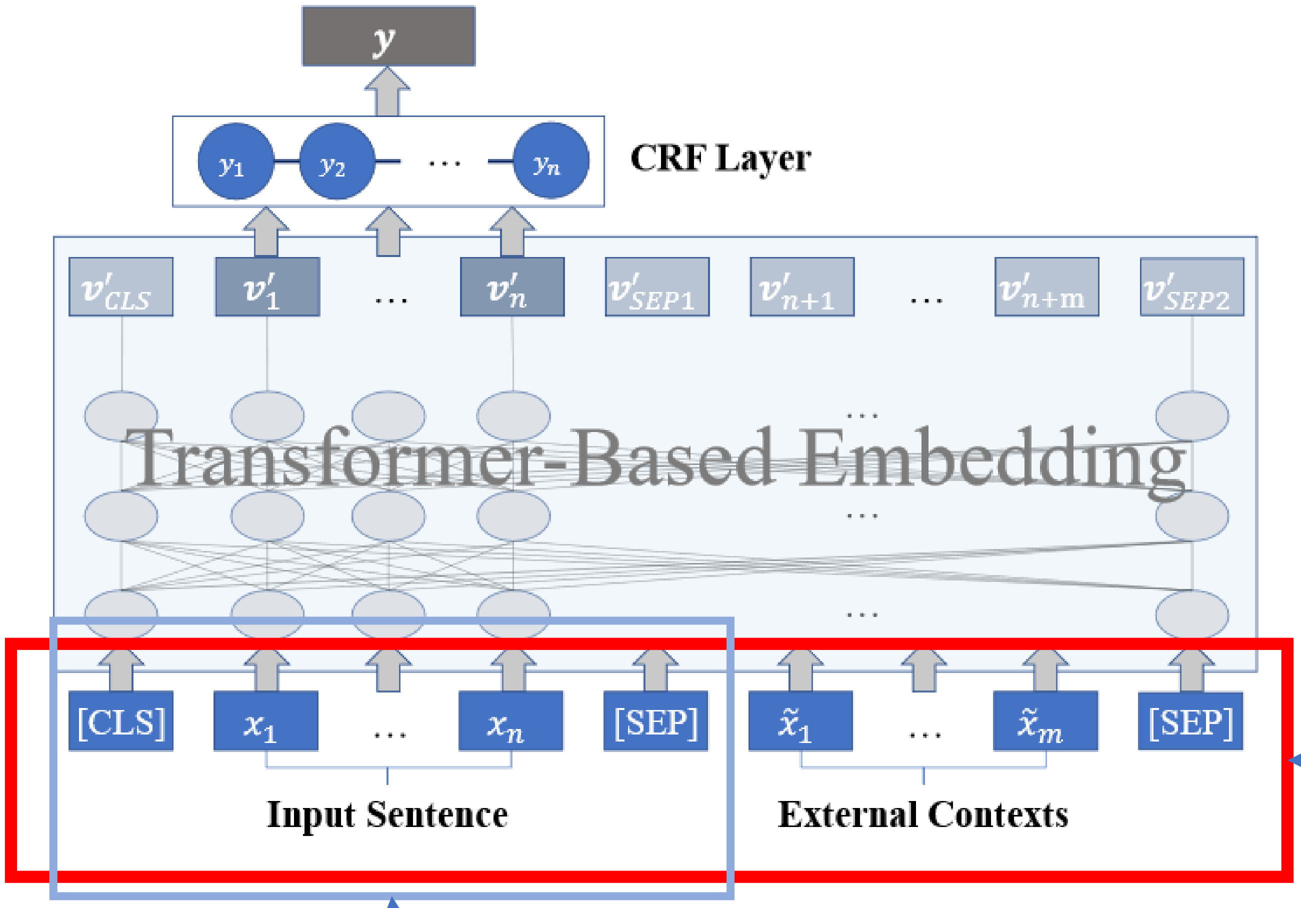


NER Model



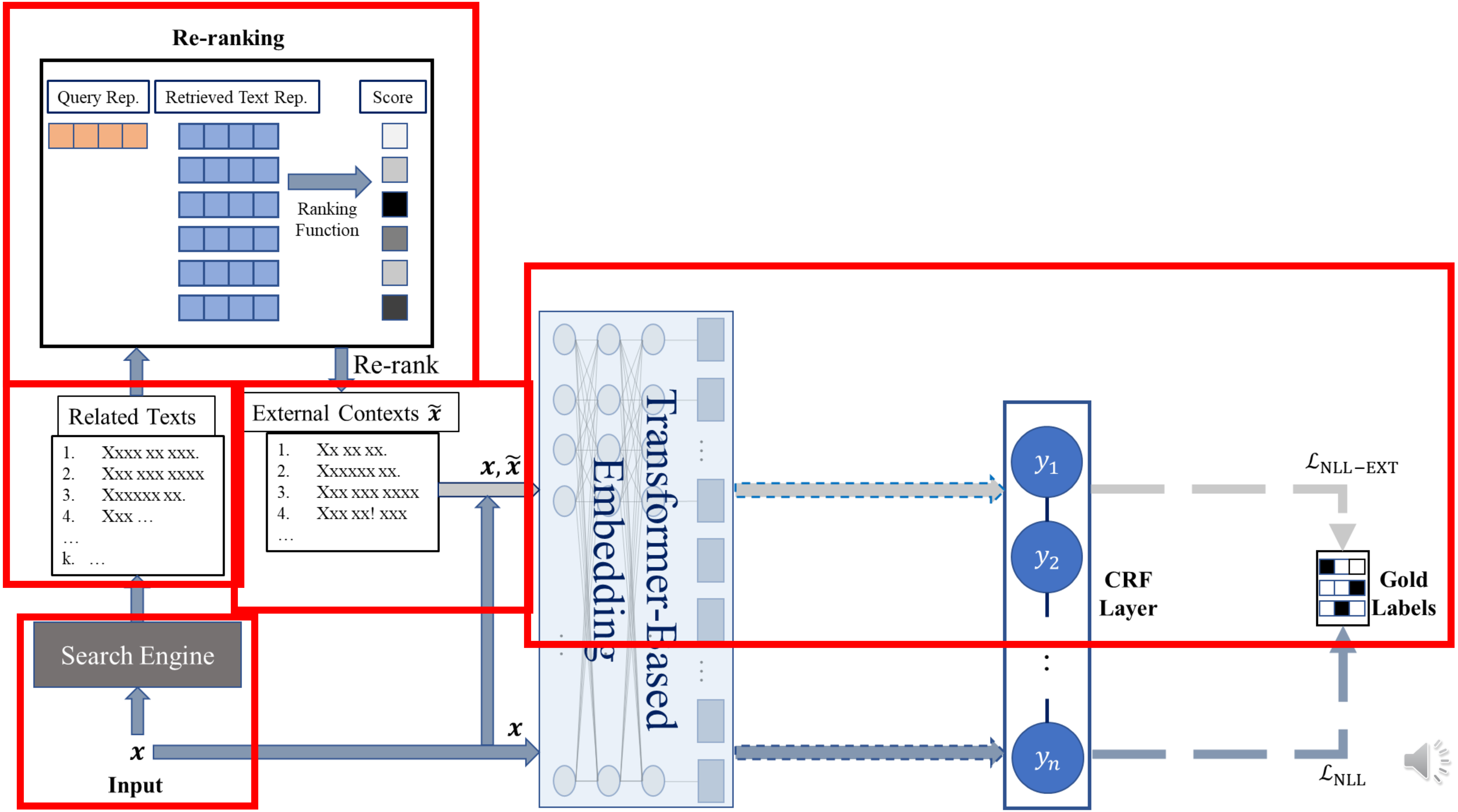
y
CRF Layer
y_1
y_2
y_n
v'_{CLS}
v'_1
v'_n
v'_{SEP1}
v'_{n+1}
v'_{n+m}
v'_{SEP2}
Transformer-Based Embedding
[CLS]
x_1
x_n
[SEP]
$\tilde{x}_1$
$\tilde{x}_m$
[SEP]
Input Sentence
External Contexts

NER Model



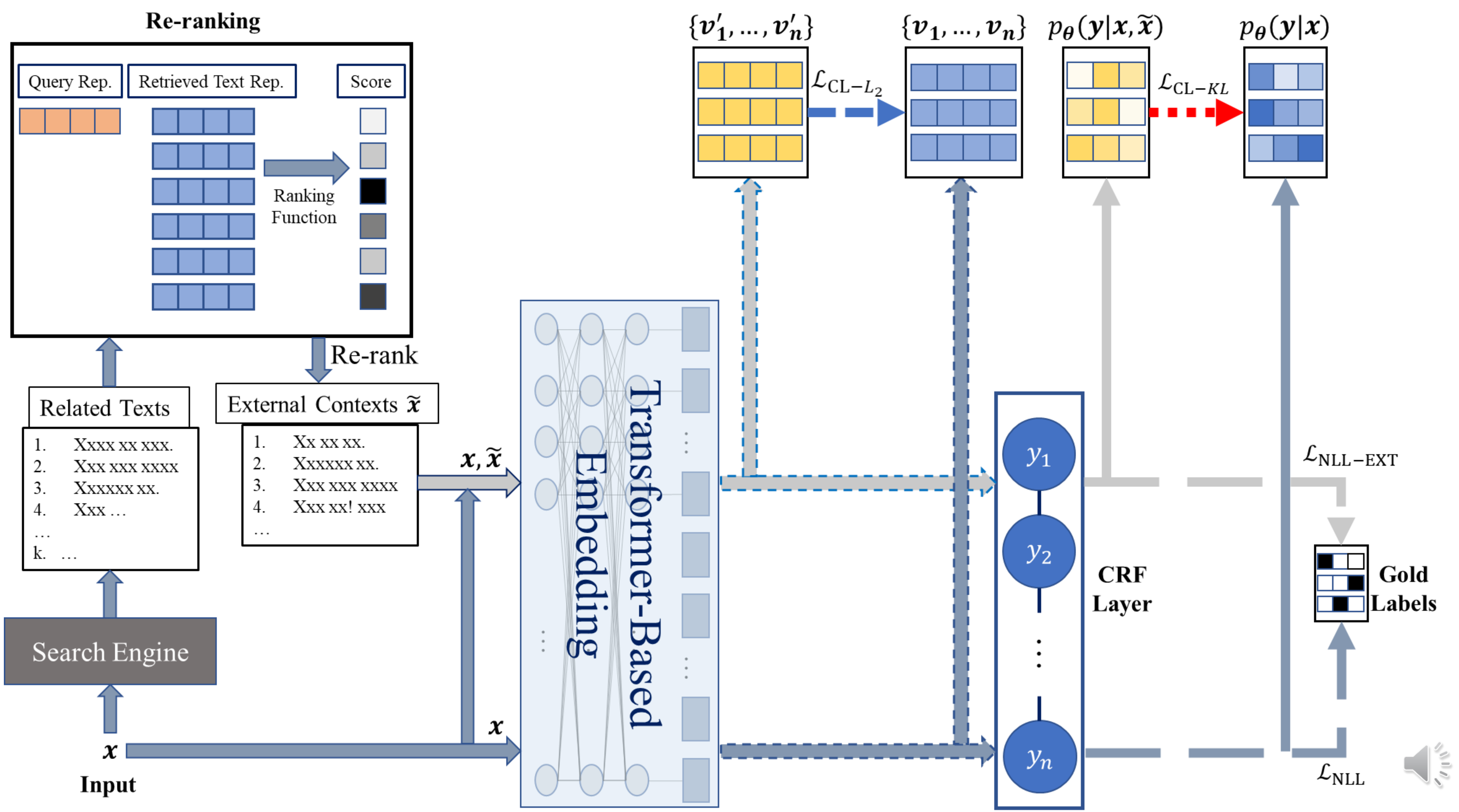
Retrieval-based Input View

Original Input View


Re-ranking
Query Rep.
Retrieved Text Rep.
Score
Ranking Function
Re-rank
Related Texts
1. Xxxx xx xxx.
2. Xxx xxx xxxx
3. Xxxxxx xx.
4. Xxx …
…
k. …
External Contexts $\tilde{x}$
1. Xx xx xx.
2. Xxxxxx xx.
3. Xxx xxx xxxx
4. Xxx xx! xxx
…
$x, \tilde{x}$
Search Engine
x
Input
x
Transformer-Based Embedding
y_1
y_2
y_n
CRF Layer
$\mathcal{L}_{\mathrm{NLL-EXT}}$
Gold Labels
$\mathcal{L}_{\mathrm{NLL}}$

Cooperative Learning

- Limitations
  - What if search engines are not available?
  - What about time-critical scenarios (no time to invoke search engines or to process the returned contexts that can be very long)?
- Solution
  - Cooperative learning (CL) between the retrieval-based input view and the original no-context input view
- Result
  - Both views are improved!



Re-ranking
Query Rep.
Retrieved Text Rep.
Score
Ranking Function
Re-rank
Related Texts
1. Xxxx xx xxx.
2. Xxx xxx xxxx
3. Xxxxxx xx.
4. Xxx …
…
k. …
External Contexts $\tilde{x}$
1. Xx xx xx.
2. Xxxxxx xx.
3. Xxx xxx xxxx
4. Xxx xx! xxx
…
Search Engine
x
Input
$x, \tilde{x}$
x
Transformer-Based Embedding
$\{v'_1, \ldots, v'_n\}$
$\{v_1, \ldots, v_n\}$
$p_\theta(y|x, \tilde{x})$
$p_\theta(y|x)$
$\mathcal{L}_{\text{CL}-L_2}$
$\mathcal{L}_{\text{CL}-KL}$
y_1
y_2
y_n
CRF Layer
$\mathcal{L}_{\text{NLL-EXT}}$
Gold Labels
$\mathcal{L}_{\text{NLL}}$

Cooperative Learning

- Objective Function:

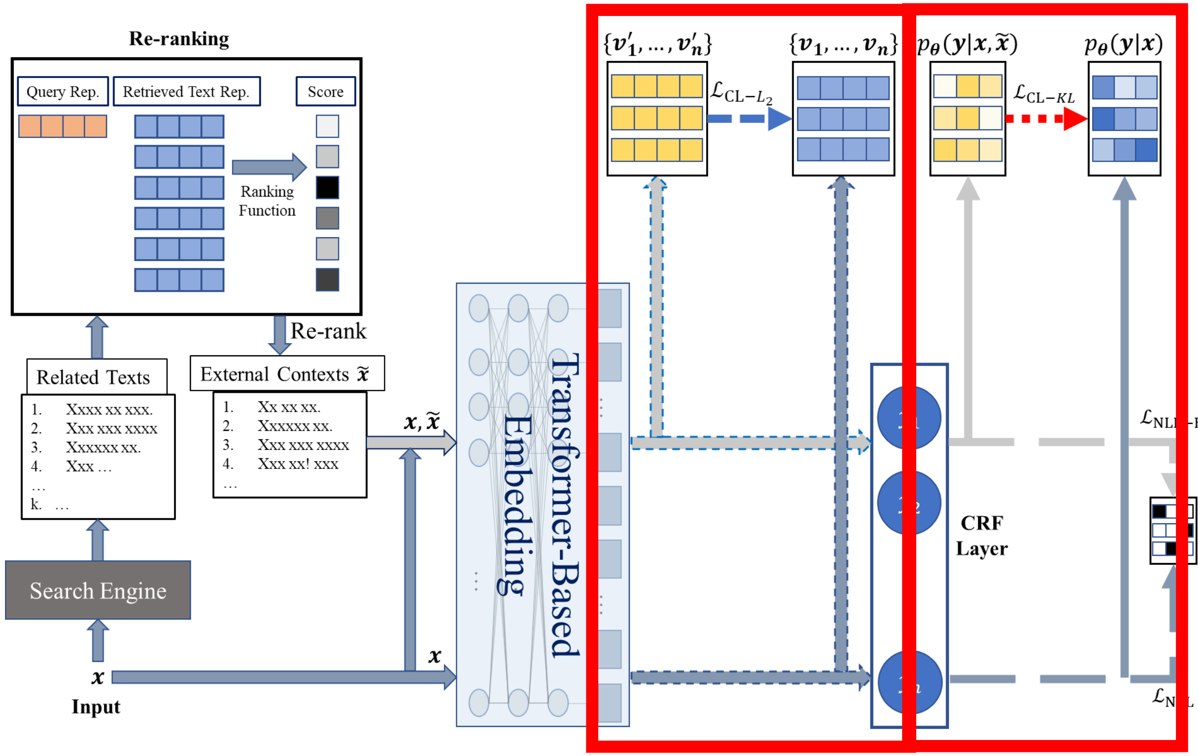
$$\mathcal{L}_{\text{CL}}(\theta) = D(h([\boldsymbol{x};\tilde{\boldsymbol{x}}]), h([\boldsymbol{x}]))$$

- Token Representations:

$$\mathcal{L}_{\text{CL-}L_2}(\theta) = \sum_{i=1}^{n} ||\boldsymbol{v}'_i - \boldsymbol{v}_i||_2^2$$

- Label Distributions:

$$\mathcal{L}_{\text{CL-KL}}(\theta)=\sum_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x})} \text{KL}(p_\theta(\boldsymbol{y}|\boldsymbol{x},\tilde{\boldsymbol{x}})||p_\theta(\boldsymbol{y}|\boldsymbol{x}))$$



Optimization

- Base Model (with original input view):

$$\mathcal{L}_{\text{NLL}}(\theta) = -\log p_\theta(\boldsymbol{y}^*|\boldsymbol{x})$$

- Training with External Contexts (with retrieval-based input view):

$$\mathcal{L}_{\text{NLL-EXT}}(\theta) = -\log p_\theta(\boldsymbol{y}^*|\boldsymbol{x}, \tilde{\boldsymbol{x}})$$

- Cooperative Learning (with both input views):

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{NLL}}(\theta) + \mathcal{L}_{\text{NLL-EXT}}(\theta) + \mathcal{L}_{\text{CL}}(\theta)$$

Results

	Social Media		News		Biomedical		E-commerce
	WNUT-16	WNUT-17	CoNLL-03	CoNLL++	BC5CDR	NCBI	
Zhou et al. (2019)	55.43	42.83	-	-	-	-	-
Nguyen et al. (2020)	52.10	56.50	-	-	-	-	-
Nie et al. (2020)	55.01	50.36	-	-	-	-	-
Baevski et al. (2019)	-	-	93.50	-	-	-	-
Wang et al. (2019)	-	-	93.43	94.28	-	-	-
Li et al. (2020)	-	-	93.33	-	-	-	-
Nooralahzadeh et al. (2019)	-	-	-	-	89.93	-	-
Bio-Flair (2019)	-	-	-	-	89.42	88.85	-
Bio-BERT (2020)	-	-	-	-	-	87.70	-
			Evaluation: w/o Context				
LUKE (2020)	54.04	55.22	92.42	93.99	89.18	87.62	77.64
w/o Context	56.04	57.86	93.03	94.20	90.52	88.65	81.47
CL-L_2	57.35†	58.68†	93.08	94.38†	90.70†	89.20†	82.43†
CL-KL	58.14†	59.33†	93.21†	94.55†	90.73†	**89.24**†	82.31†
			Evaluation: w/ Context				
w/ Context	57.43†	60.20†	93.27†	94.56†	90.76†	89.01†	83.15†
CL-L_2	58.61†	60.26†	93.47†	94.62†	**90.99**†	89.22†	83.87†
CL-KL	**58.98**†	**60.45**†	**93.56**†	**94.81**†	90.93†	88.96†	**83.99**†

Results

	Social Media		News		Biomedical		E-commerce
	WNUT-16	WNUT-17	CoNLL-03	CoNLL++	BC5CDR	NCBI	
Zhou et al. (2019)	55.43	42.83	-	-	-	-	-
Nguyen et al. (2020)	52.10	56.50	-	-	-	-	-
Nie et al. (2020)	55.01	50.36	-	-	-	-	-
Baevski et al. (2019)	-	-	93.50	-	-	-	-
Wang et al. (2019)	-	-	93.43	94.28	-	-	-
Li et al. (2020)	-	-	93.33	-	-	-	-
Nooralahzadeh et al. (2019)	-	-	-	-	89.93	-	-
Bio-Flair (2019)	-	-	-	-	89.42	88.85	-
Bio-BERT (2020)	-	-	-	-	-	87.70	-
Evaluation: W/O CONTEXT							
LUKE (2020)	54.04	55.22	92.42	93.99	89.18	87.62	77.64
W/O CONTEXT	56.04	57.86	93.03	94.20	90.52	88.65	81.47
CL-L_2	57.35†	58.68†	93.08	94.38†	90.70†	89.20†	82.43†
CL-KL	58.14†	59.33†	93.21†	94.55†	90.73†	**89.24**†	82.31†
Evaluation: W/ CONTEXT							
W/ CONTEXT	57.43†	60.20†	93.27†	94.56†	90.76†	89.01†	83.15†
CL-L_2	58.61†	60.26†	93.47†	94.62†	**90.99**†	89.22†	83.87†
CL-KL	**58.98**†	**60.45**†	**93.56**†	**94.81**†	90.93†	88.96†	**83.99**†

Results

	Social Media		News		Biomedical		E-commerce
	WNUT-16	WNUT-17	CoNLL-03	CoNLL++	BC5CDR	NCBI	
Zhou et al. (2019)	55.43	42.83	-	-	-	-	-
Nguyen et al. (2020)	52.10	56.50	-	-	-	-	-
Nie et al. (2020)	55.01	50.36	-	-	-	-	-
Baevski et al. (2019)	-	-	93.50	-	-	-	-
Wang et al. (2019)	-	-	93.43	94.28	-	-	-
Li et al. (2020)	-	-	93.33	-	-	-	-
Nooralahzadeh et al. (2019)	-	-	-	-	89.93	-	-
Bio-Flair (2019)	-	-	-	-	89.42	88.85	-
Bio-BERT (2020)	-	-	-	-	-	87.70	-
Evaluation: W/O CONTEXT							
LUKE (2020)	54.04	55.22	92.42	93.99	89.18	87.62	77.64
W/O CONTEXT	56.04	57.86	93.03	94.20	90.52	88.65	81.47
CL-L_2	57.35†	58.68†	93.08	94.38†	90.70†	89.20†	82.43†
CL-KL	58.14†	59.33†	93.21†	94.55†	90.73†	**89.24**†	82.31†
Evaluation: W/ CONTEXT							
W/ CONTEXT	57.43†	60.20†	93.27†	94.56†	90.76†	89.01†	83.15†
CL-L_2	58.61†	60.26†	93.47†	94.62†	**90.99**†	89.22†	83.87†
CL-KL	**58.98**†	**60.45**†	**93.56**†	**94.81**†	90.93†	88.96†	**83.99**†

Results

	Social Media		News		Biomedical		E-commerce
	WNUT-16	WNUT-17	CoNLL-03	CoNLL++	BC5CDR	NCBI	
Zhou et al. (2019)	55.43	42.83	-	-	-	-	-
Nguyen et al. (2020)	52.10	56.50	-	-	-	-	-
Nie et al. (2020)	55.01	50.36	-	-	-	-	-
Baevski et al. (2019)	-	-	93.50	-	-	-	-
Wang et al. (2019)	-	-	93.43	94.28	-	-	-
Li et al. (2020)	-	-	93.33	-	-	-	-
Nooralahzadeh et al. (2019)	-	-	-	-	89.93	-	-
Bio-Flair (2019)	-	-	-	-	89.42	88.85	-
Bio-BERT (2020)	-	-	-	-	-	87.70	-
Evaluation: W/O CONTEXT							
LUKE (2020)	54.04	55.22	92.42	93.99	89.18	87.62	77.64
W/O CONTEXT	56.04	57.86	93.03	94.20	90.52	88.65	81.47
CL-L_2	57.35†	58.68†	93.08	94.38†	90.70†	89.20†	82.43†
CL-KL	58.14†	59.33†	93.21†	94.55†	90.73†	**89.24**†	82.31†
Evaluation: W/ CONTEXT							
W/ CONTEXT	57.43†	60.20†	93.27†	94.56†	90.76†	89.01†	83.15†
CL-L_2	58.61†	60.26†	93.47†	94.62†	**90.99**†	89.22†	83.87†
CL-KL	**58.98**†	**60.45**†	**93.56**†	**94.81**†	90.93†	88.96†	**83.99**†

Results

	Social Media		News		Biomedical		E-commerce
	WNUT-16	WNUT-17	CoNLL-03	CoNLL++	BC5CDR	NCBI	
Zhou et al. (2019)	55.43	42.83	-	-	-	-	-
Nguyen et al. (2020)	52.10	56.50	-	-	-	-	-
Nie et al. (2020)	55.01	50.36	-	-	-	-	-
Baevski et al. (2019)	-	-	93.50	-	-	-	-
Wang et al. (2019)	-	-	93.43	94.28	-	-	-
Li et al. (2020)	-	-	93.33	-	-	-	-
Nooralahzadeh et al. (2019)	-	-	-	-	89.93	-	-
Bio-Flair (2019)	-	-	-	-	89.42	88.85	-
Bio-BERT (2020)	-	-	-	-	-	87.70	-
Evaluation: W/O CONTEXT							
LUKE (2020)	54.04	55.22	92.42	93.99	89.18	87.62	77.64
W/O CONTEXT	56.04	57.86	93.03	94.20	90.52	88.65	81.47
CL-L_2	57.35†	58.68†	93.08	94.38†	90.70†	89.20†	82.43†
CL-KL	58.14†	59.33†	93.21†	94.55†	90.73†	**89.24**†	82.31†
Evaluation: W/ CONTEXT							
W/ CONTEXT	57.43†	60.20†	93.27†	94.56†	90.76†	89.01†	83.15†
CL-L_2	58.61†	60.26†	93.47†	94.62†	**90.99**†	89.22†	83.87†
CL-KL	**58.98**†	**60.45**†	**93.56**†	**94.81**†	90.93†	88.96†	**83.99**†

Transfer Learning

	Evaluation **Science and Technology**	
Approach	W/O CONTEXT	W/ CONTEXT
Jia et al. (2019)	73.59	-
W/O CONTEXT	75.87	75.74
W/ CONTEXT	75.72	75.94
CL-L_2	76.16	76.10
CL-KL	**76.37**	**76.38**

Semi-supervised Learning on E-Commerce

	Evaluation	
Approach	W/O CONTEXT	W/ CONTEXT
CL-L_2	82.43	83.87
CL-KL	82.31	83.99
CL–L_2+SEMI	**82.88**†	83.92
CL-KL+SEMI	82.58†	**84.10**

How the Context Quality Affects Accuracy?

	WNUT-17
w/ Context (Ours)	**60.20**
w/o Context	57.86
w/ Context (Dataset)	57.21
w/ Context (Generated)	57.71
w/ Context (Random Retrieved)	57.53
w/ Context (Random Data)	47.69

Table 6: A comparison among different contexts types.

Conclusion

- Improving NER models' accuracy by retrieving related contexts
- Cooperative Learning improves the robustness of the models when no external contexts are available
- A state-of-the-art accuracy on the NER datasets over 5 domains