Finding Cycles in Graph: A Unified Approach for Various NER Tasks

Liang Qiao[1], Pengfei Li[2], Ting Jin[3], and Xi Li[1]

Zhejiang University, China
Xi'an Jiaotong University, China
Hainan University, China

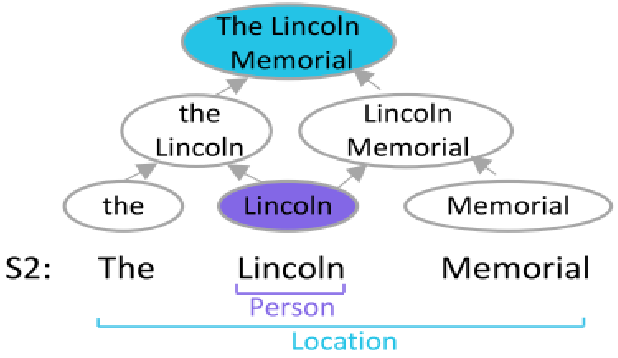
CONTENT

1. Problem
2. Method
3. Experiments
4. Contact

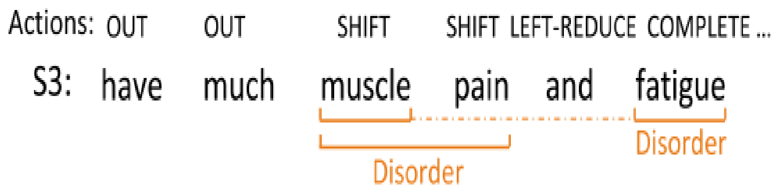
1. Problem

Classical NER (Named Entity Recognition) Paradigms

(a) Sequence labelling methods for ***flat NER***



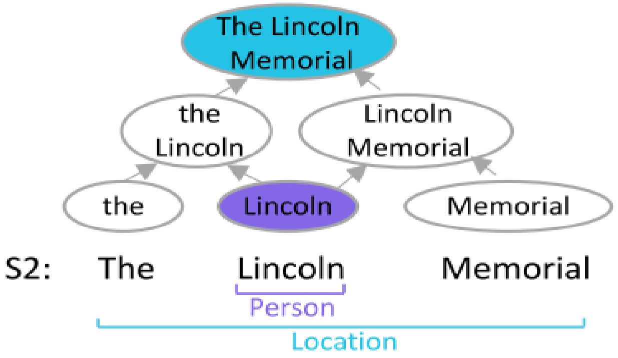
(b) Span-based methods for ***overlapped NER***



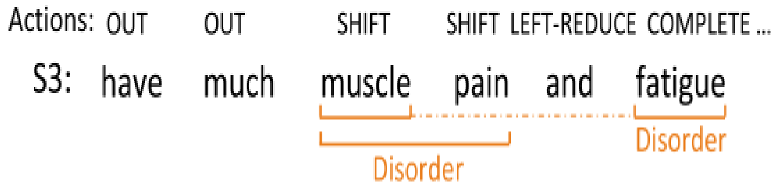
(c) Transition/hypergraph-based methods for ***discontinuous NER***

Classical NER (Named Entity Recognition) Paradigms

(a) Sequence labelling methods for ***flat NER***



(b) Span-based methods for ***overlapped NER***



(c) Transition/hypergraph-based methods for ***discontinuous NER***

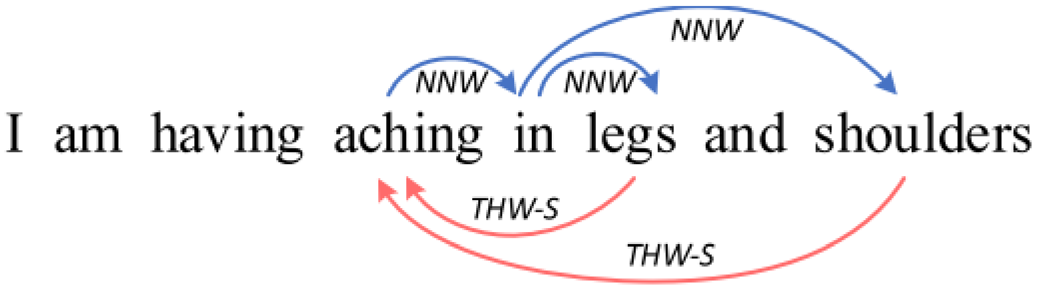
Unified NER Paradigms

S1: Barack Obama <Person> US <Location>

S2: The Lincoln Memorial <Location> Lincoln <Person>

S2: muscle pain < Disorder > muscle fatigue <Disorder>

generative paradigms (Yan et al., 2021)

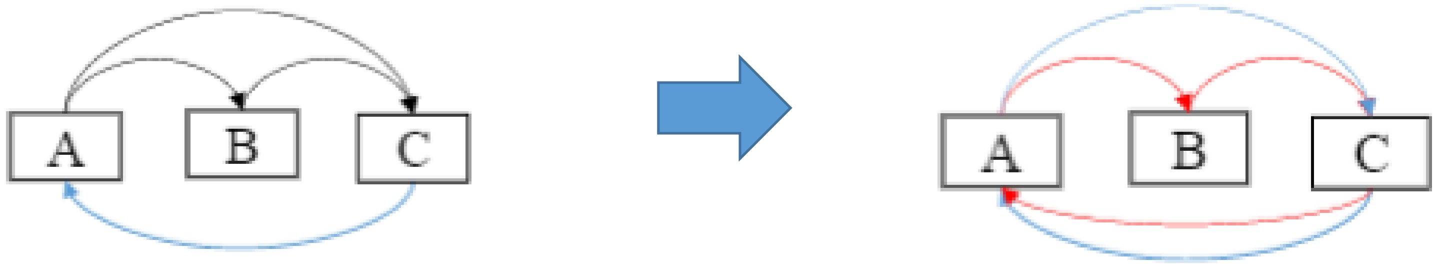


word-word relation paradigms (Li et al., 2022)

Problems in Word-Word Relation Paradigms

P1: The paradigm cannot theoretically support all kind of entities.

Q: What if "ABC" and "AC" belong to different types?



THW-*

All of the edges should be classified in multiple types

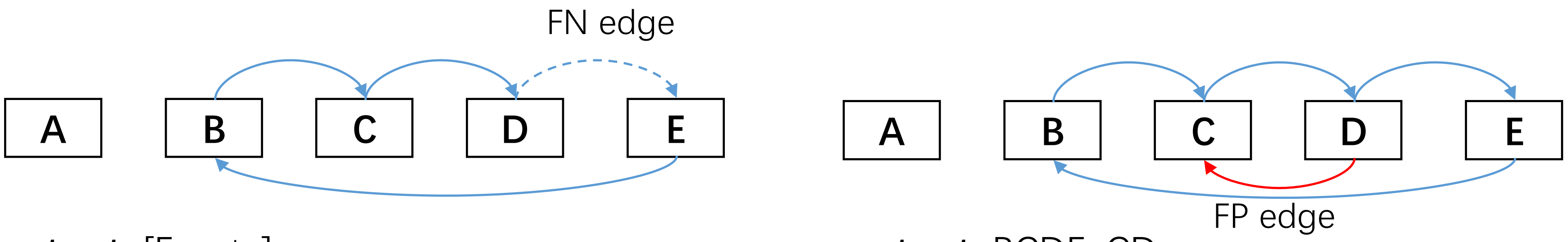
Q: What if there is an entity of "CBA" (An extreme example) ?

All of the edges could be considered equally

Problems in Word-Word Relation Paradigms

P2: The learning process is error-prone

Q: What if one of the edge being misclassified?



output: [Empty]

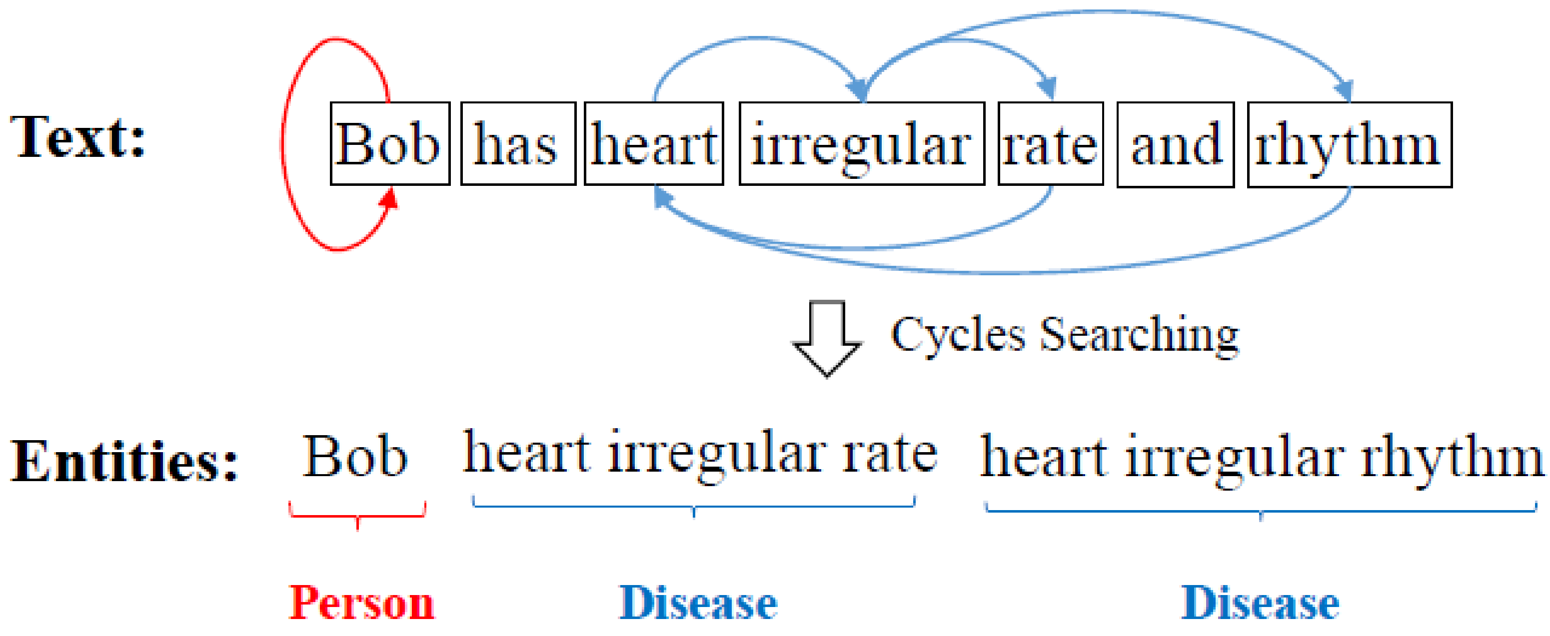
output: BCDE, CD

Small loss but big impact

Entities should be considered as a whole.

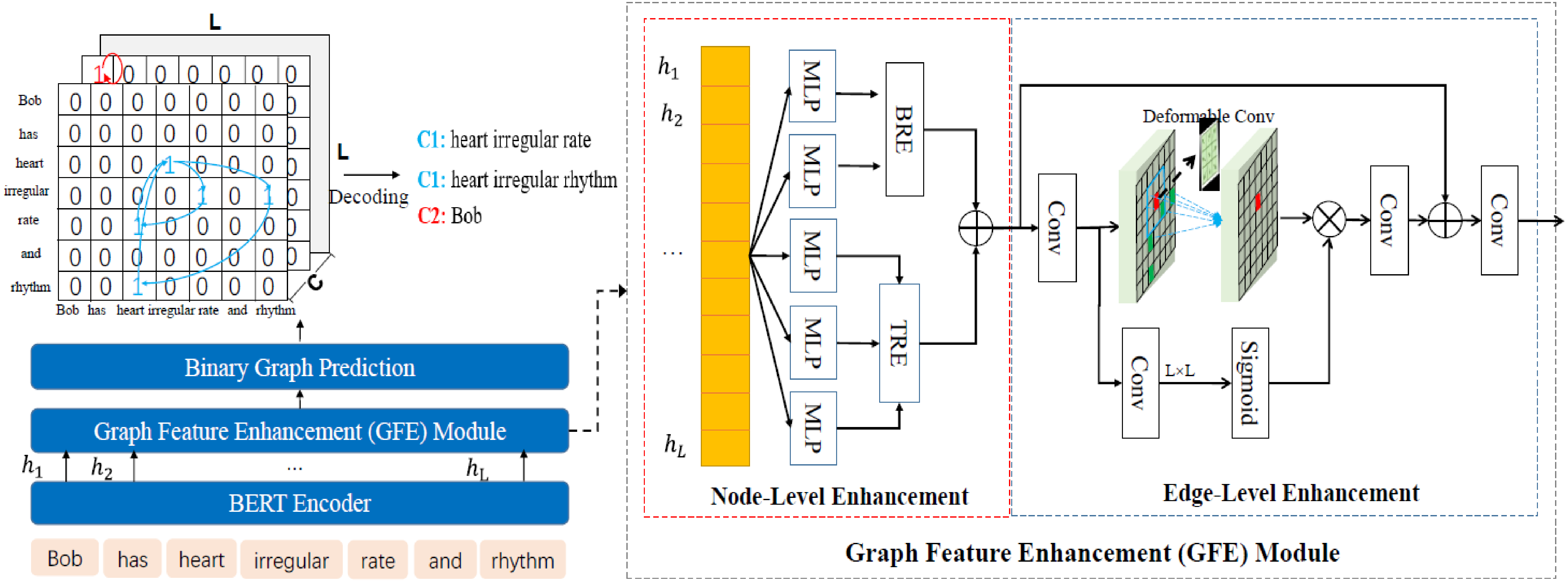
2. Method

Treat Entities as Cycles



Core motivation： **Enhance the complete cycles formation.**

Framework of CycleNER



Graph Feature Enhancement Module

Node-Level

- **input**: L x C
- **output:** L x L x C

Binary Relation Extraction

$$\mathcal{F}^{Q}, \mathcal{F}^{K} = MLP^{Q,K}(H),$$

$$\mathcal{F}_{ij}^{A} = (\mathcal{F}_{i}^{Q} \oplus \mathcal{F}_{j}^{K})W^{A},$$

$$\mathcal{F}_{ij}^{D} = ((\mathcal{F}_{i}^{Q} R_i) \otimes (\mathcal{F}_{j}^{K} R_j))W^{D},$$

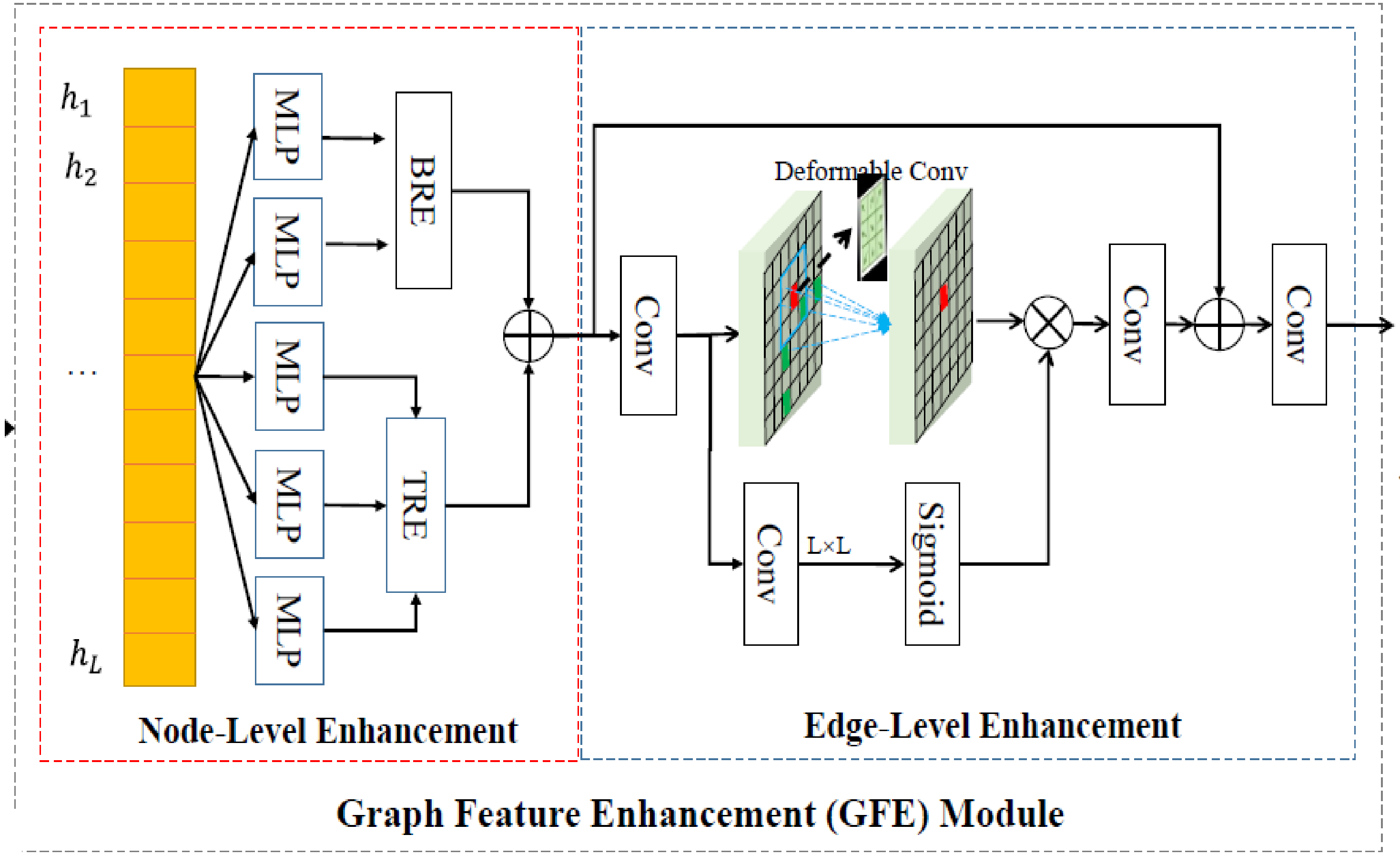
$$\mathcal{F}^{BRE} = \mathcal{F}^{A} + \mathcal{F}^{D}.$$

Ternary Relation Extraction

$$\mathcal{F}^{X}, \mathcal{F}^{Y}, \mathcal{F}^{Z} = MLP^{X,Y,Z}(H),$$

$$\mathcal{T}_{ijk} = \mathcal{F}_{i}^{X\top} \mathcal{F}_{j}^{Y\top} W^{T} \mathcal{F}_{k}^{Z},$$

$$\mathcal{F}^{TRE} = GlobalMaxPooling(\mathcal{T}).$$



Edge-Level

- **input**: L x L x C
- **output:** L x L x M

$$\mathcal{F}_{edge} = conv_{1\times1}(\mathcal{F})$$

$$\mathcal{F}_{dcn} = dconv_{3\times3}(dconv_{3\times3}(\mathcal{F}_{edge}))$$

$$\mathcal{F}_{att} = \mathcal{F}_{dcn} \otimes (sigmoid(conv_{3\times3}(\mathcal{F}_{edge}))),$$

$$\mathcal{F}_{out} = conv_{1\times1}(conv_{1\times1}(\mathcal{F}_{att}) + \mathcal{F}).$$

Optimized with Cycle Loss

BCE loss:

$$L_{bce} = -\frac{1}{NCL^2}\sum_{i=1}^{N}\sum_{c=1}^{C}\sum_{m=1}^{L}\sum_{n=1}^{L} y_{icmn}\log \hat{y}_{icmn},$$

Cycle loss:

$$s_1 = diagonal(A_G).$$

$$\mathcal{T}_k = (\mathcal{T}_{k-1} - M(diagonal(\mathcal{T}_{k-1})))\mathcal{T}_1,$$

$$\mathcal{T}_1 = (\mathcal{A}_G - M(diagonal(\mathcal{A}_G))),$$

$$s_k = diagonal(\mathcal{T}_k).$$

$$L_{cycle} = \sum_{k=1}^{K}\sum_{i=1}^{L}(\hat{s}_{k,i} - s_{k,i})^2,$$

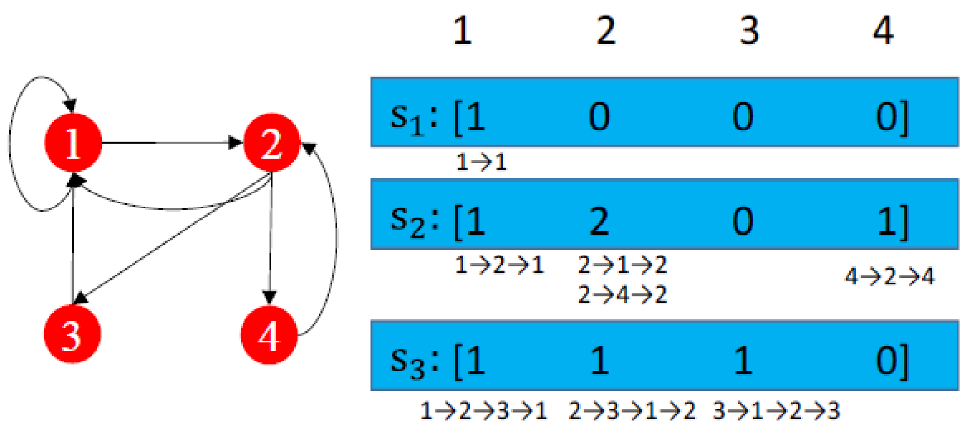
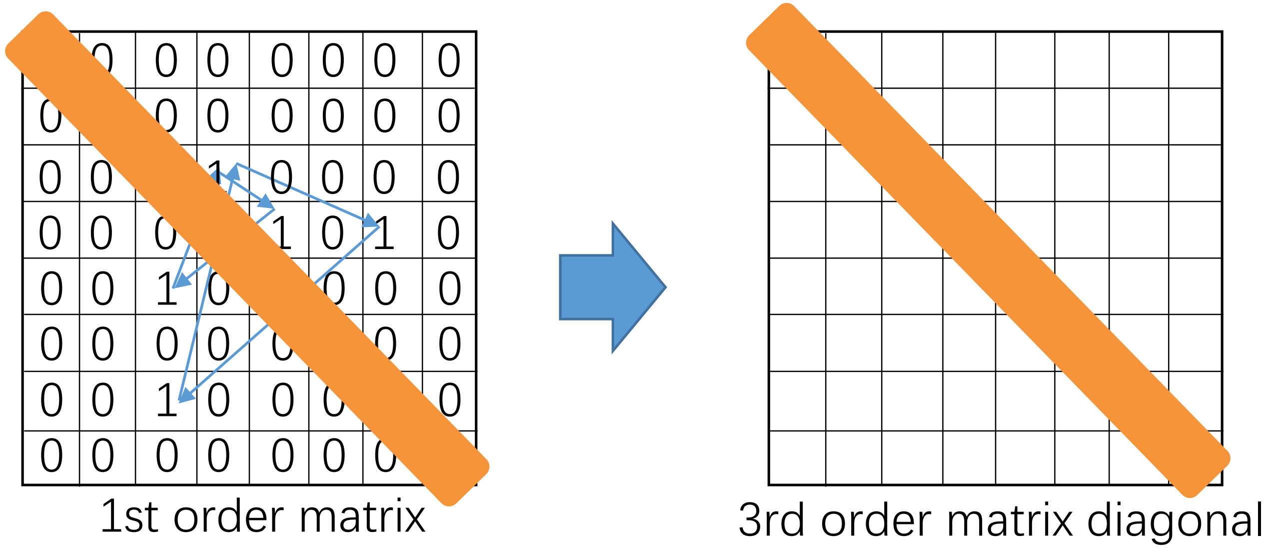


Fig. 3: An example of s_k that represents the vector that contains the number of cycles (exclude nested cycles) with the length k.



3. Experiment

Flat NER datasets results:

Types	Methods	Pretrained-models	CoNLL2003			OntoNotes		
			P	R	F1	P	R	F1
Sequence-labeling	(Peters et al., 2018) [35]	[ELMO]	-	-	92.22	-	-	-
Sequence-labeling	(Devlin et al., 2019) [15]	[BERT-Large]	-	-	92.80	-	-	-
Span-based	(Yu et al., 2020) [7]†	[BERT-Large]	92.85	92.15	92.50	89.92	89.74	89.83
Span-based	(Li et al.,2020) [8]]†	[BERT-Large]	92.47	93.27	92.87	91.34	88.39	89.84
Hypergraph-based	(Wang and Lu, 2018) [10]	[Glove]	-	-	90.50	-	-	-
Unified	(Yan et al., 2021) [3]	[BART-Large]	92.61	**93.87**	**93.24**	89.99	90.77	90.38
Unified	(Li et al.,2022) [14]	[BERT-Large]	92.71	93.44	93.07	90.03	**90.97**	**90.50**
Unified	CycleNER(ours)	[BERT-Large]	**93.50**	91.91	92.70	**91.69**	88.68	90.16

TABLE I: Results for flat NER datasets. Results with "†" are reported from [3].

Overlapped NER datasets results:

Types	Methods	Pretrained-models	ACE2004			ACE2005			GENIA		
			P	R	F1	P	R	F1	P	R	F1
Sequence-labeling	(Straková et al., 2019) [36]	[BERT-Large]	-	-	84.33	-	-	83.42	-	-	78.20
Sequence-labeling	(Shibuya and Hovy, 2020) [20]	[BERT-Large]	-	-	-	83.30	84.69	83.99	77.46	76.65	77.05
Span-based	(Yu et al., 2020) [7]†	[BERT-Large]	85.42	85.92	85.67	84.50	84.72	84.61	79.43	78.32	78.87
Span-based	(Li et al., 2020) [8]†	[BERT-Large]	85.83	85.77	85.80	85.01	84.13	84.57	**81.25**	76.36	78.72
Span-based	(Shen et al., 2021) [25]∗	[Glove&BERT-Large]	87.20	87.26	87.23	86.24	86.54	86.39	78.47	79.19	78.83
Span-based	(Tan et al., 2021) [37]∗	[Glove&BERT-Large]	87.86	85.63	86.73	86.88	86.15	86.51	80.73	77.20	78.93
Span-based	(Li et al., 2021a) [4]∗	[BERT-Large]	86.58	86.10	86.34	83.11	85.39	84.23	78.88	77.31	78.09
Hypergraph-based	(Wang and Lu, 2018) [10]	[Glove]	78.00	72.40	75.10	76.80	72.30	74.50	77.00	73.30	75.10
Unified	(Yan et al., 2021) [3]	[BART-Large]	87.27	86.41	86.84	83.16	86.38	84.74	78.87	**79.60**	**79.23**
Unified	(Li et al., 2022) [14]∗	[BERT-Large]	87.90	87.08	87.49	84.78	**88.04**	86.38	80.55	77.32	78.90
Unified	CycleNER(ours)	[BERT-Large]	**88.45**	**87.99**	**88.22**	**87.48**	86.47	**86.97**	79.23	77.48	78.35

TABLE II: Results for overlapped NER datasets. Results with "†" are reported from [3], and "∗" means our re-implemented result with the BERT-Large pretrained model.

Models	ACE2004	ACE2005	GENIA
(Yan et al., 2021) [3]	70.64/-/-	79.69/-/55.0	80.34/-/52.7
(Dai et al., 2020) [13]	69.0/65.4/37.9	77.7/62.9/52.5	79.6/63.1/49.2
(Li et al., 2022) [14]∗	87.49/87.34/82.41	86.38/87.05/77.62	**78.90**/75.45/38.52
CycleNER (ours)	**88.22/88.53/83.25**	**86.97/87.79/78.51**	78.35/**75.50/38.83**

TABLE III: Performance on Overlapped Entities, '/' separates the overall results, the result of sentences including overlapped entity, and the result only considers overlapped entities. '∗' means our re-implemented result with the BERT-Large pre-trained model.

Discontinuous NER datasets results:

Types	Methods	Pretrained-models	CADEC			ShARe 13			ShARe14		
			P	R	F1	P	R	F1	P	R	F1
Sequence-labeling	(Tang et al., 2018) [16]	[Glove]	67.80	64.99	66.36	-	-	-	-	-	-
Span-based	(Wang et al., 2021) [9]∗	[BERT-Large]	70.50	72.50	71.50	83.25	76.46	79.71	78.20	83.65	80.83
Span-based	(Li et al., 2021a) [4]∗	[BERT-Large]	70.10	69.05	69.57	82.50	76.79	79.54	79.23	81.10	80.15
Hypergraph-based	(Wang and Lu, 2019) [11]	[Word-Embedding]	72.10	48.40	58.00	83.80	60.40	70.30	79.10	70.70	74.70
Others	(Dai et al., 2020) [13]	[ELMO]	68.90	69.00	69.00	80.50	75.00	77.70	78.10	81.20	79.60
Others	(Fei et al., 2021) [47]∗	[BERT-Large]	73.11	70.25	71.65	83.68	76.23	79.78	78.23	82.62	80.37
Unified	(Yan et al., 2021) [3]	[BART-Large]	70.08	71.21	70.64	82.09	**77.42**	79.69	77.20	**83.75**	80.34
Unified	(Li et al., 2022) [14]∗	[BERT-Large]	72.02	70.28	71.14	82.63	76.75	79.58	79.96	80.14	80.05
Unified	CycleNER(ours)	[BERT-Large]	**73.49**	**71.49**	**72.48**	**84.96**	76.35	**80.43**	**80.30**	82.13	**81.20**

TABLE IV: Results for discontinuous NER datasets. '∗' means our re-implemented with the BERT-Large pretrained model.

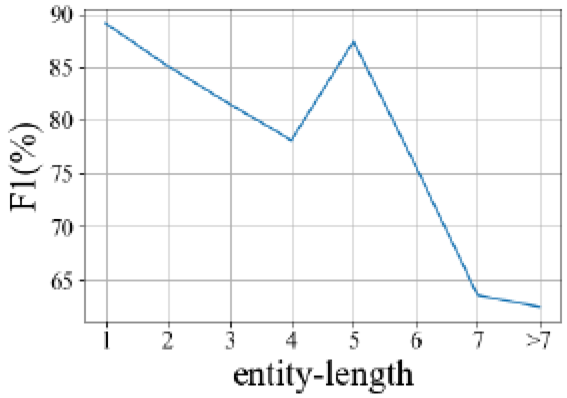
Models	CADEC	ShARe13	ShARe14
(Yan et al., 2021) [3]	70.64/-/-	79.69/-/55.0	80.34/-/52.7
(Dai et al., 2020) [13]	69.0/65.4/37.9	77.7/62.9/52.5	79.6/63.1/49.2
(Li et al., 2022) [14]∗	71.1/67.8/45.1	79.6/65.8/56.7	80.1/65.5/49.9
CycleNER (ours)	**72.4/71.3/48.8**	**80.4/66.6/57.5**	**81.2/69.3/53.9**

TABLE V: Performance on Discontinuous Entities, '/' separates the overall results, the result of sentences with at least one discontinuous entity, and the result only considers discontinuous entities.'∗' means our re-implemented result with the BERT-Large pretrained model.

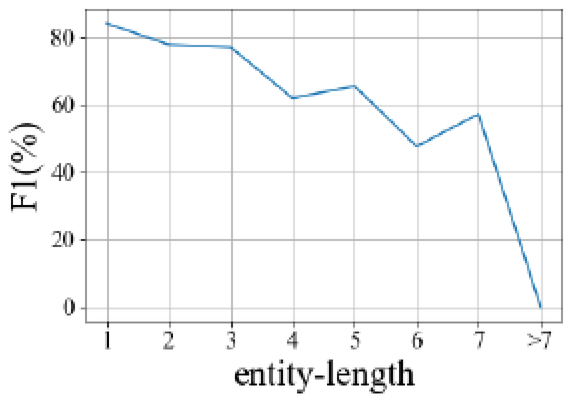
Ablation Studies & Statistic Analysis:

	CADEC	ShARe13	ShARe14
CycleNER (ours)	**72.48**	**80.43**	**81.20**
-BRE	14.49(↓57.99)	19.08(↓61.35)	10.39 (↓70.63)
-TRE	71.41(↓1.07)	79.90(↓0.53)	80.78(↓0.24)
-Edge-level Enhancement	71.85(↓0.63)	79.84(↓0.59)	80.64(↓0.38)
-Cycle Loss	70.90(↓1.58)	79.72(↓0.71)	80.50(↓0.50)

TABLE VI: Ablation results on discontinuous NER datasets, '-' denotes remove the component alone.



(a) ACE2005



(b) ShARe13

Fig. 5: Impact of entity length on ACE2005 and ShARe13

4. Contact

Email to: qiaoliang@zju.edu.cn

THANKS !