武漢大学
WUHAN UNIVERSITY

Unified Named Entity Recognition as Word-Word Relation Classification

Jingye Li,[1,*] Hao Fei,[1,*] Jiang Liu,[1] Shengqiong Wu,[1]
Meishan Zhang,[2] Chong Teng,[1] Donghong Ji,[1] Fei Li[1,†]

[1] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, China
[2] Institute of Computing and Intelligence, Harbin Institute of Technology (Shenzhen), China
{theodorelee, hao.fei, liujiang, whuwsq, tengchong, dhji, lifei_csnlp}@whu.edu.cn, mason.zms@gmail.com

Background

- Name Entity Recognition

At issue is the liability of a Pennsylvania radio station under the federal wiretap statute.

O O O O O O O B I I O O O O O

- Two assumptions
  - Entities should not be overlapped.
  - Entities should be continuous.

Irregular Named Entities

- Overlapped Entities

Pennsylvania radio station

Pennsylvania

At issue is the liability of a Pennsylvania radio station under the federal wiretap statute.

O O O O O O O B I I O O O O O

- Discontinuous Entities

mitral leaflets thickened

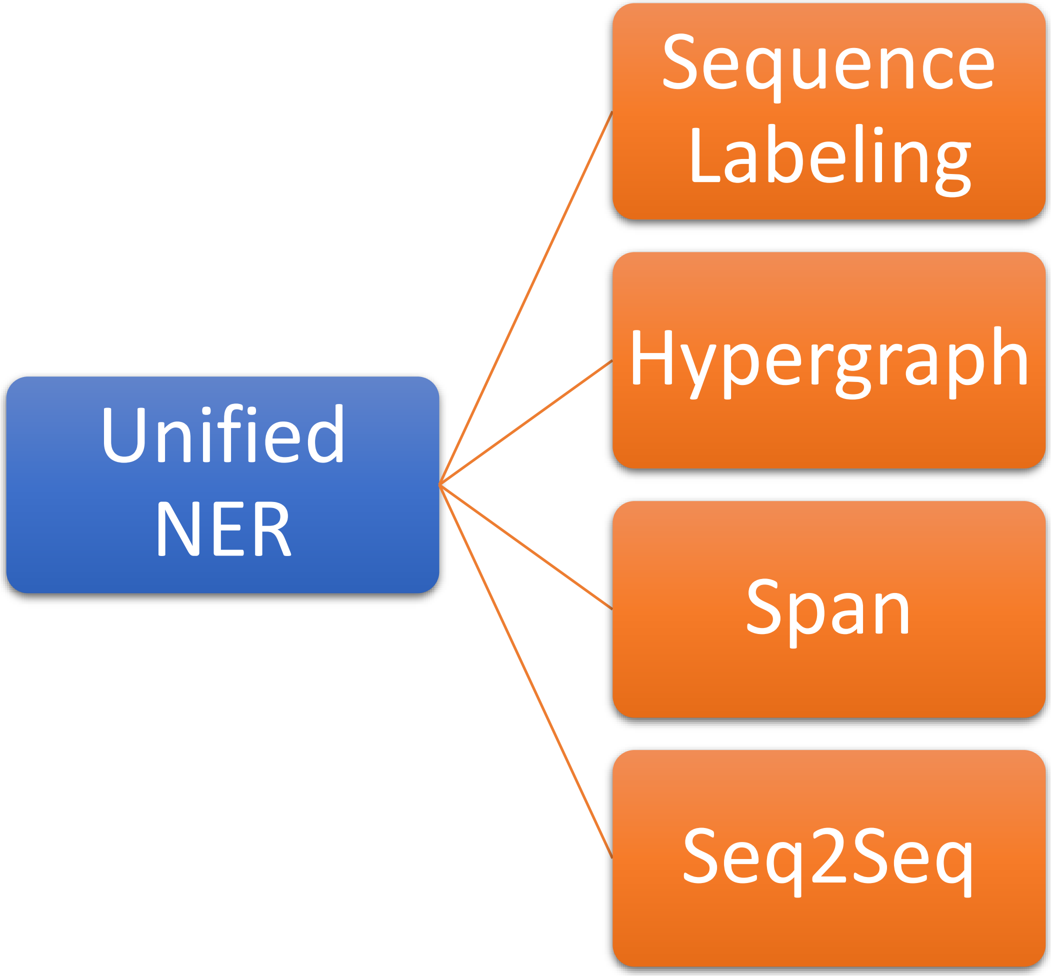
The mitral valve leaflets are mildly thickened

O B O B/I O O B/I

Unified NER

- One model that is able to recognize
  - Regular entities (non-overlapped, continuous)
  - Overlapped entities (including nested entities)
  - Discontinuous entities

Related Work



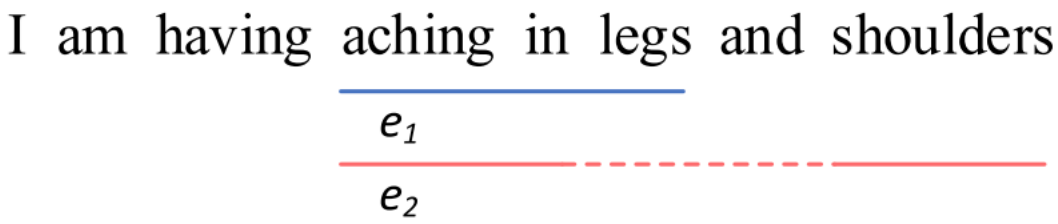
Tang, Buzhou, et al. "Recognizing disjoint clinical concepts in clinical text using machine learning-based methods." *AMIA annual symposium proceedings*. Vol. 2015. American Medical Informatics Association, 2015.

Aldrian Obaja Muis and Wei Lu. 2016. Learning to recognize discontiguous entities. In Proceedings ofthe 2016 Conference on EMNLP, pages 75–84.

Luan, Y.; Wadden, D.; He, L.; Shah, A.; Ostendorf, M.; and Hajishirzi, H. 2019. A general framework for information extraction using dynamic span graphs. In Proceedings of the NAACL, 3036–3046.

Yan, H.; Gui, T.; Dai, J.; Guo, Q.; Zhang, Z.; and Qiu, X. 2021. A Unified Generative Framework for Various NER Subtasks. In Proceedings ofthe ACL-IJCNLP, 5808–5822.

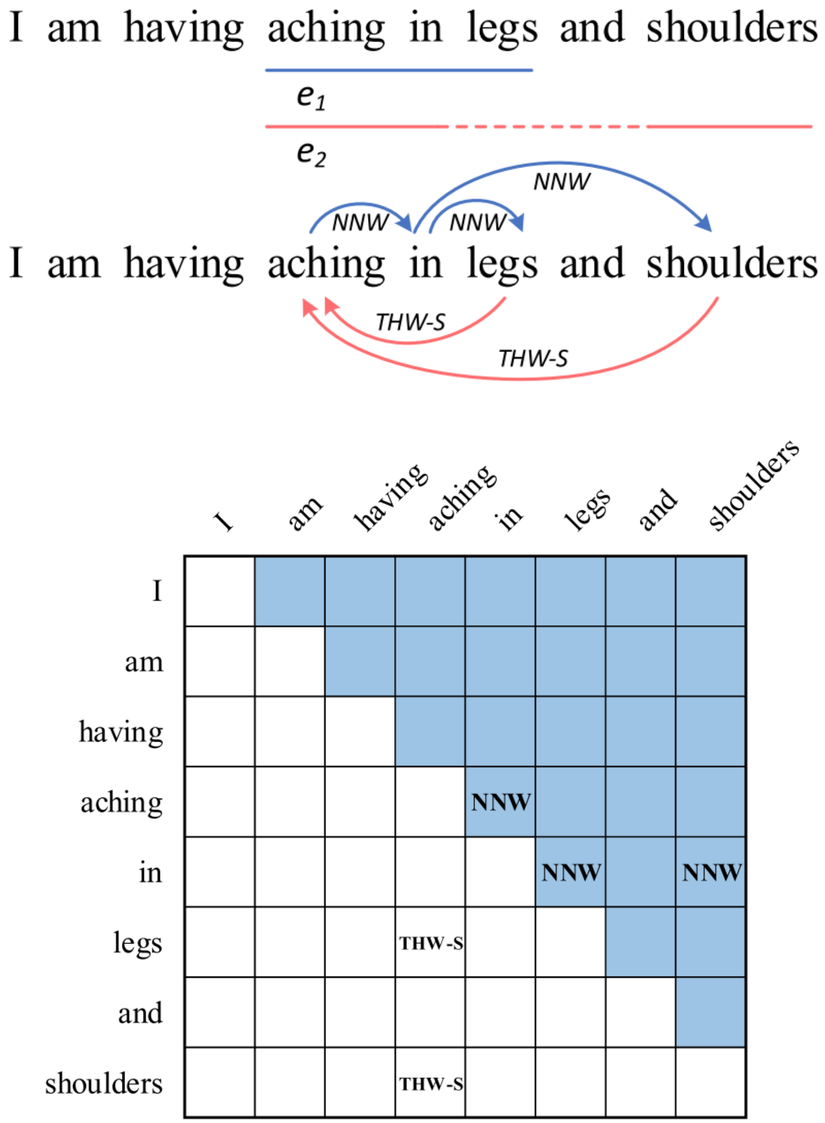
Motivation



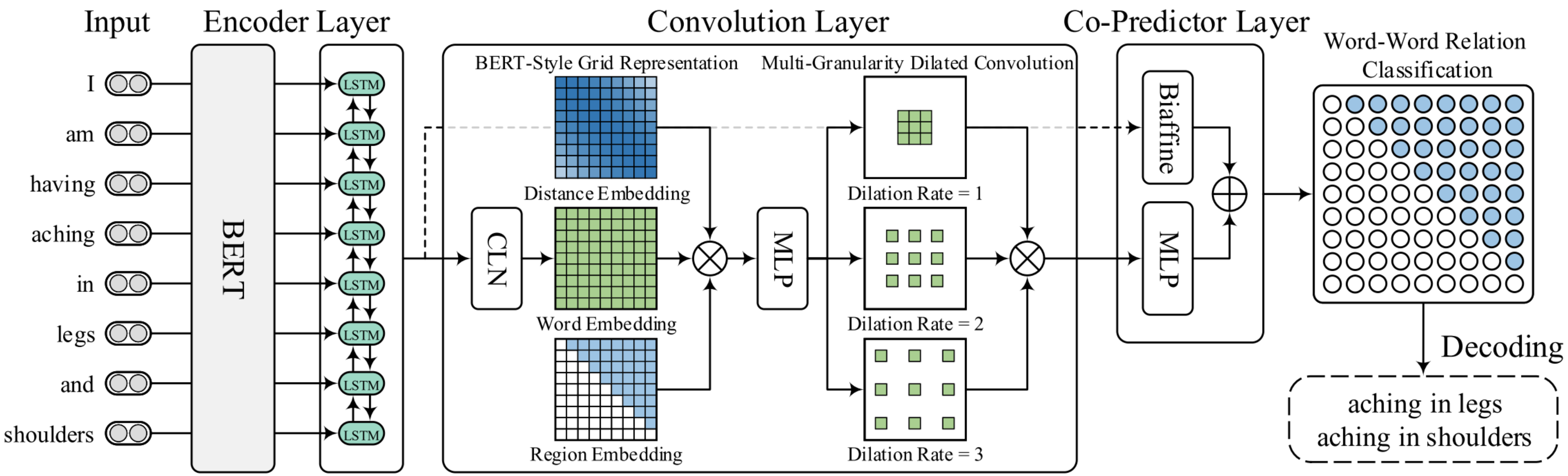
- Most of the existing work has paid the major focus on how to identify the entity boundary.
- They neglect the neighboring relations between entity words, which are important for unified NER.
- Such adjacency correlations describe the semantic connectivity between the partial text segments, which plays the key role for the overlapping and discontinuous ones.

Our Labeling Scheme --- W²NER

- Two kinds of word-word relations
- NNW
  - Next-Neighboring-Word
  - the NNW relation indicates that the word pair belongs to an entity mention
- THW-*
  - Tail-Head-Word-*
  - the THW relation indicates that the word in certain row of the grid is the tail of an entity mention

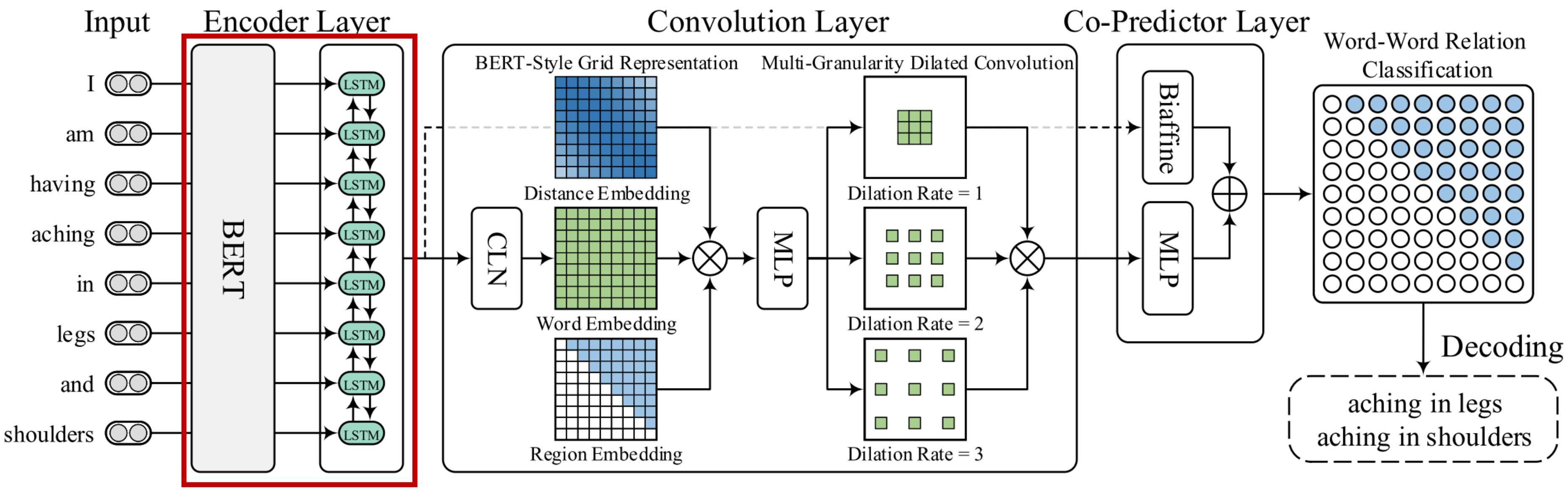


Model



- Encoder Layer
- Convolution Layer
- Co-Predictor Layer

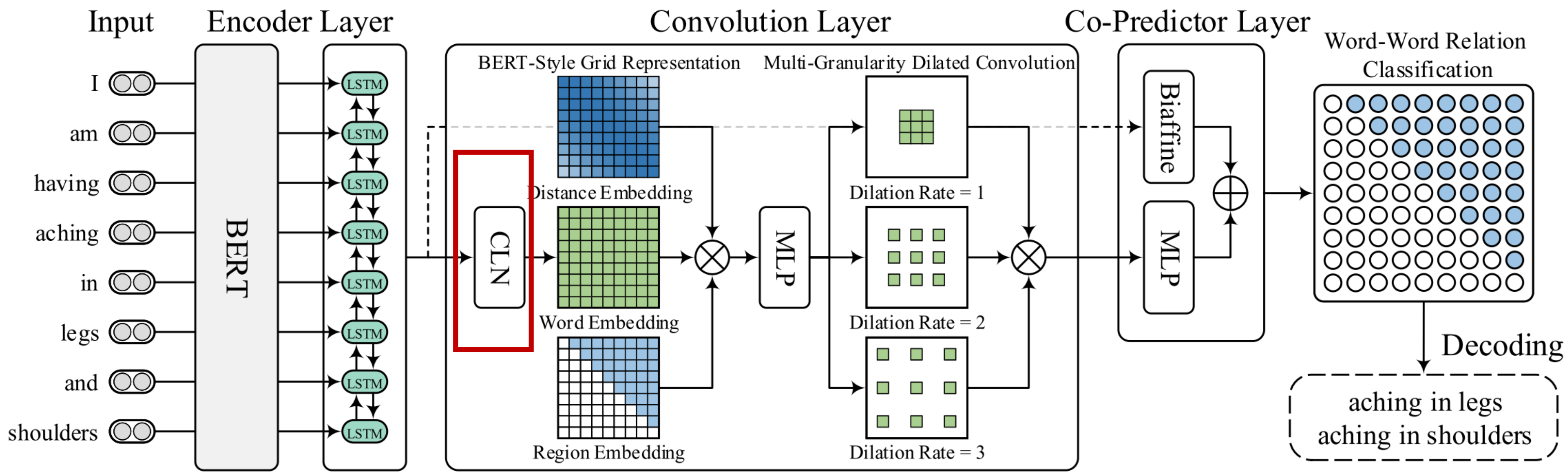
Model



- Encoder Layer
  - BERT
    - max pooling (from word piece representations to word representations)
  - bi-directional LSTM

$$\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_N\} \in \mathbb{R}^{N \times d_h}$$

Model



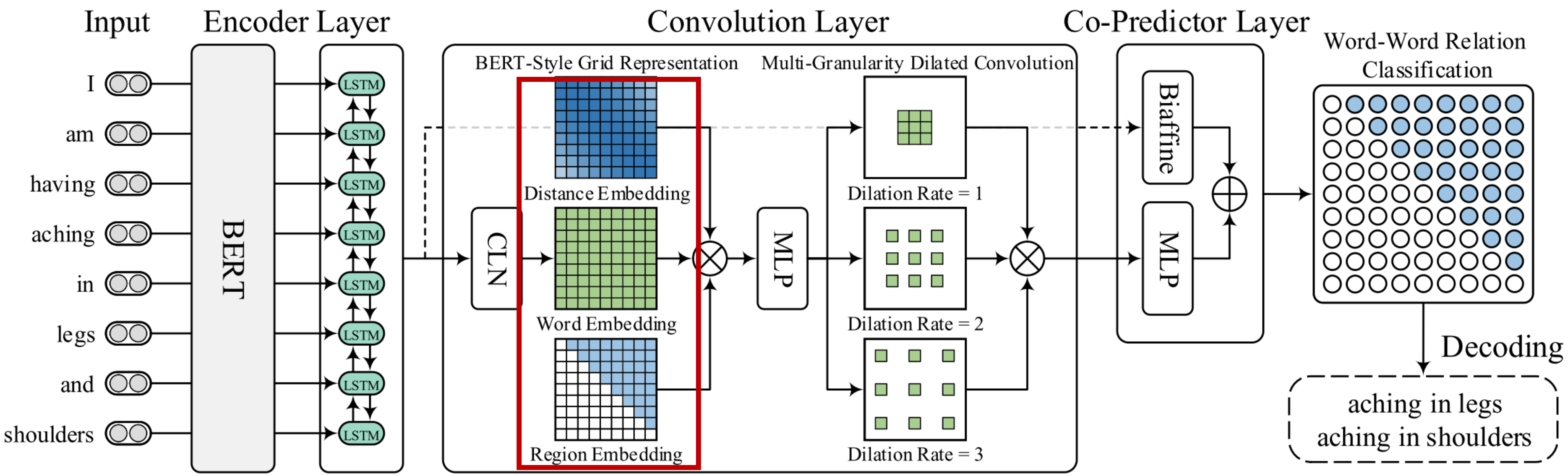
- Convolution Layer
  - Conditional Layer Normalization

$$\mathbf{V}_{ij} = \text{CLN}(\mathbf{h}_i, \mathbf{h}_j) = \gamma_{ij} \odot (\frac{\mathbf{h}_j - \mu}{\sigma}) + \lambda_{ij}$$

$$\gamma_{ij} = \mathbf{W}_\alpha \mathbf{h}_i + \mathbf{b}_\alpha \qquad \lambda_{ij} = \mathbf{W}_\beta \mathbf{h}_i + \mathbf{b}_\beta$$

$$\mu = \frac{1}{d_h}\sum_{k=1}^{d_h} h_{jk}, \quad \sigma = \sqrt{\frac{1}{d_h}\sum_{k=1}^{d_h}(h_{jk} - \mu)^2}$$

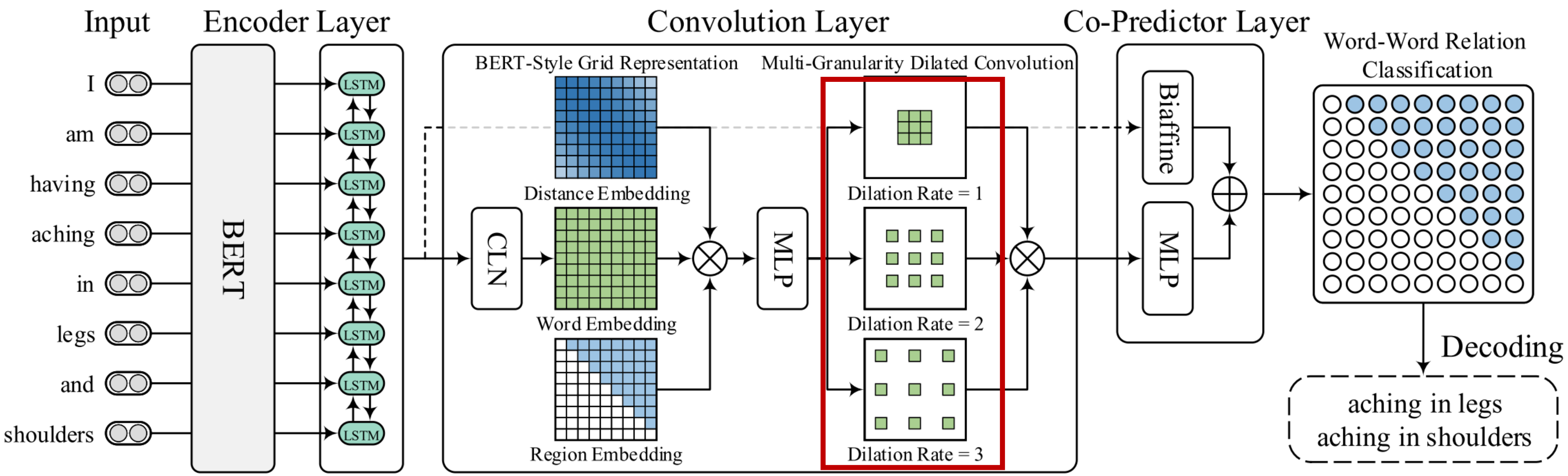
Model



- Convolution Layer
  - BERT-Style Grid Representation

$$\mathbf{C} = \mathrm{MLP}_1([\mathbf{V}; \mathbf{E}^d; \mathbf{E}^t])$$

Model

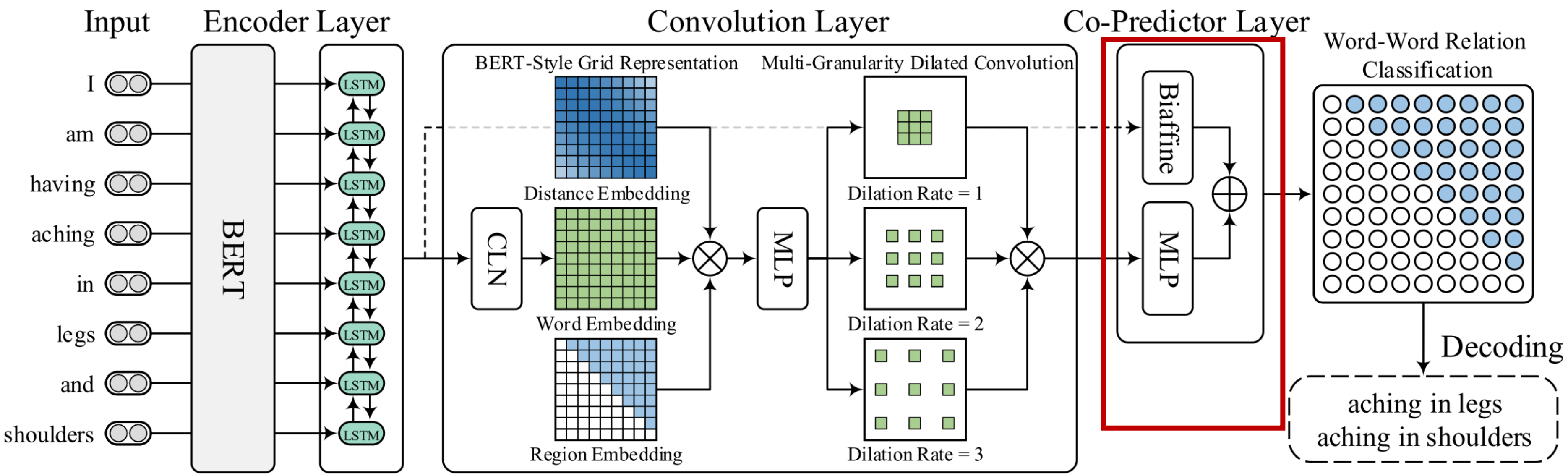


- Convolution Layer
  - Multi-Granularity Dilated Convolution

$$\mathbf{Q}^l = \sigma(\mathrm{DConv}_l(\mathbf{C}))$$

$$[\mathbf{Q}^1, \mathbf{Q}^2, \mathbf{Q}^3] \in \mathbb{R}^{N \times N \times 3d_c}$$

Model



- Co-Predictor Layer
  - MLP Predictor

$$\mathbf{y}''_{ij} = \mathrm{MLP}(\mathbf{Q}_{ij})$$

  - Biaffine Predictor

$$\mathbf{s}_i = \mathrm{MLP}_2(\mathbf{h}_i) \ ,$$
$$\mathbf{o}_j = \mathrm{MLP}_3(\mathbf{h}_j) \ ,$$
$$\mathbf{y}'_{ij} = \mathbf{s}_i{}^{\top}\mathbf{U}\mathbf{o}_j + \mathbf{W}[\mathbf{s}_i;\mathbf{o}_j] + \mathbf{b}$$

  - Combination

$$\mathbf{y}_{ij} = \mathrm{Softmax}(\mathbf{y}'_{ij} + \mathbf{y}''_{ij})$$

Training

- minimize the negative log-likelihood losses

$$X = \{x_1, x_2, ..., x_N\}$$

$$\mathcal{L} = -\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{r=1}^{|\mathcal{R}|} \hat{\mathbf{y}}_{ij}^{r} \mathbf{log} \mathbf{y}_{ij}^{r}$$

Decoding

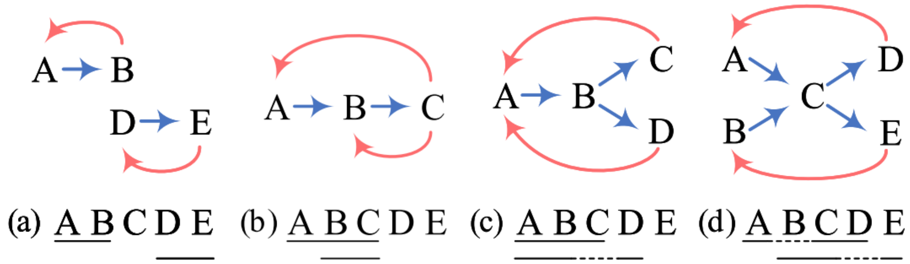


Figure 4: Four decoding cases for the word sequence "ABCDE". (a) "AB" and "DE" are flat entities. (b) The flat entity "BC" is nested in "ABC". (c) The entity "ABC" is overlapped with a discontinuous entity "ABD". (d) Two discontinuous entities "ACD" and "BCE" are overlapped. The blue and red arrows indicate NNW and THW relations.

- The predictions of our model are the words and their relations, which can be considered as a directional word graph.
- The decoding object is to find certain paths from one word to anther word in the graph using NNW relations.
- Each path corresponds to an entity mention.
- Besides the type and boundary identification for NER, THW relations can also be used as auxiliary information for disambiguation.

Experiments --- 14 Datasets

- Flat NER
  - CoNLL-2003 (English)
  - OntoNotes 5.0 (English)
  - OntoNotes 4.0 (Chinese)
  - MSRA (Chinese)
  - Weibo (Chinese)
  - Resume (Chinese)
- Overlapped NER
  - ACE 2004 (English, Chinese)
  - ACE 2005 (English, Chinese)
  - GENIA (English)
- Discontinuous NER
  - CADEC (English)
  - ShARe13 (English)
  - ShARe14 (English)

Results

		CoNLL2003			OntoNotes 5.0		
		P	R	F1	P	R	F1
• **Sequence Labeling**	Lample et al. (2016)	-	-	90.94	-	-	-
	Strubell et al. (2017)	-	-	90.65	-	-	86.84
• **Span-based**	Yu et al. (2020) †	**92.91**	92.13	92.52	90.01	89.77	89.89
	Shen et al. (2021)	92.13	**93.73**	92.94	-	-	-
• **Hypergraph-based**	Wang and Lu (2018)	-	-	90.50	-	-	-
• **Seq2Seq**	Straková et al. (2019)	-	-	92.98	-	-	-
	Yan et al. (2021) †	92.56	93.56	93.05	89.62	90.92	90.27
	W^2NER (ours)	92.71	93.44	**93.07**	**90.03**	**90.97**	**90.50**

	OntoNotes 4.0			MSRA			Resume			Weibo		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Zhang and Yang (2018)	76.35	71.56	73.88	93.57	92.79	93.18	94.81	94.11	94.46	53.04	62.25	58.79
Yan et al. (2019)	-	-	72.43	-	-	92.74	-	-	95.00	-	-	58.17
Gui et al. (2019)	76.40	72.60	74.45	94.50	92.93	93.71	95.37	94.84	95.11	57.14	66.67	59.92
Li et al. (2020b)	-	-	81.82	-	-	96.09	-	-	95.86	-	-	68.55
Ma et al. (2020)	**83.41**	82.21	82.81	95.75	95.10	95.42	96.08	96.13	96.11	**70.94**	67.02	70.50
W^2NER (ours)	82.31	**83.36**	**83.08**	**96.12**	**96.08**	**96.10**	**96.96**	**96.35**	**96.65**	70.84	**73.87**	**72.32**

		ACE2004			ACE2005			GENIA		
		P	R	F1	P	R	F1	P	R	F1
• **Sequence Labeling**	Ju et al. (2018)	-	-	-	74.20	70.30	72.20	78.50	71.30	74.70
• **Span-based**	Wang et al. (2020)	86.08	86.48	86.28	83.95	85.39	84.66	79.45	78.94	79.19
	Yu et al. (2020)	87.30	86.00	86.70	85.20	85.60	85.40	81.80	79.30	80.50
	Shen et al. (2021)	**87.44**	87.38	87.41	**86.09**	87.27	86.67	80.19	**80.89**	80.54
• **Hypergraph-based**	Wang and Lu (2018)	78.00	72.40	75.10	76.80	72.30	74.50	77.00	73.30	75.10
• **Seq2Seq**	Straková et al. (2019)	-	-	84.33	-	-	83.42	-	-	78.20
	Yan et al. (2021)	87.27	86.41	86.84	83.16	86.38	84.74	78.87	79.60	79.23
	W^2NER (ours)	87.33	**87.71**	**87.52**	85.03	**88.62**	**86.79**	**83.10**	79.76	**81.39**

	ACE2004	ACE2005
Yu et al. (2020) ⋆	87.35	88.39
Shen et al. (2021) ⋆	87.47	88.21
W^2NER (ours)	**88.00**	**88.81**

		CADEC			ShARe13			ShARe14		
		P	R	F1	P	R	F1	P	R	F1
• **Sequence Labeling**	Tang et al. (2018)	67.80	64.99	66.36	-	-	-	-	-	-
• **Span-based**	Li et al. (2021a)	-	-	69.90	-	-	82.50	-	-	-
• **Hypergraph-based**	Wang and Lu (2019)	72.10	48.40	58.00	83.80	60.40	70.30	79.10	70.70	74.70
• **Seq2Seq**	Yan et al. (2021)	70.08	71.21	70.64	82.09	77.42	79.69	77.20	83.75	80.34
	Fei et al. (2021)	**75.50**	71.80	72.40	**87.90**	77.20	80.30	-	-	-
• **Others**	Dai et al. (2020)	68.90	69.00	69.00	80.50	75.00	77.70	78.10	81.20	79.60
	Wang et al. (2021)	70.50	**72.50**	71.50	84.30	78.20	81.20	78.20	**84.70**	81.30
	W^2NER (ours)	74.09	72.35	**73.21**	85.57	**79.68**	**82.52**	**79.88**	83.71	**81.75**

Codes available at https://github.com/ljynlp/W2NER.git

If any question, please contact

theodorelee@whu.edu.cn or lifei_csnlp@whu.edu.cn