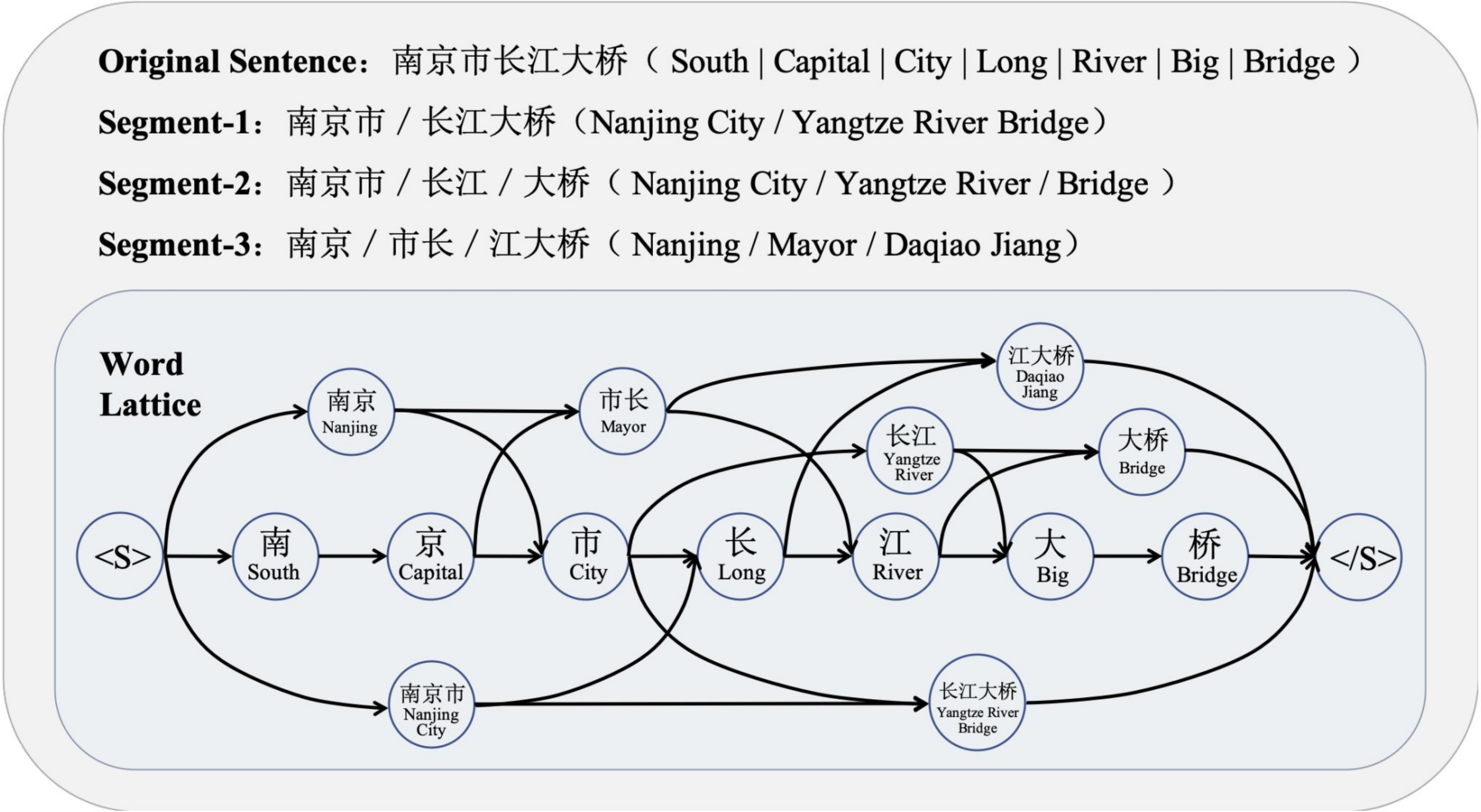
Neural Graph Matching Networks for Chinese Short Text Matching

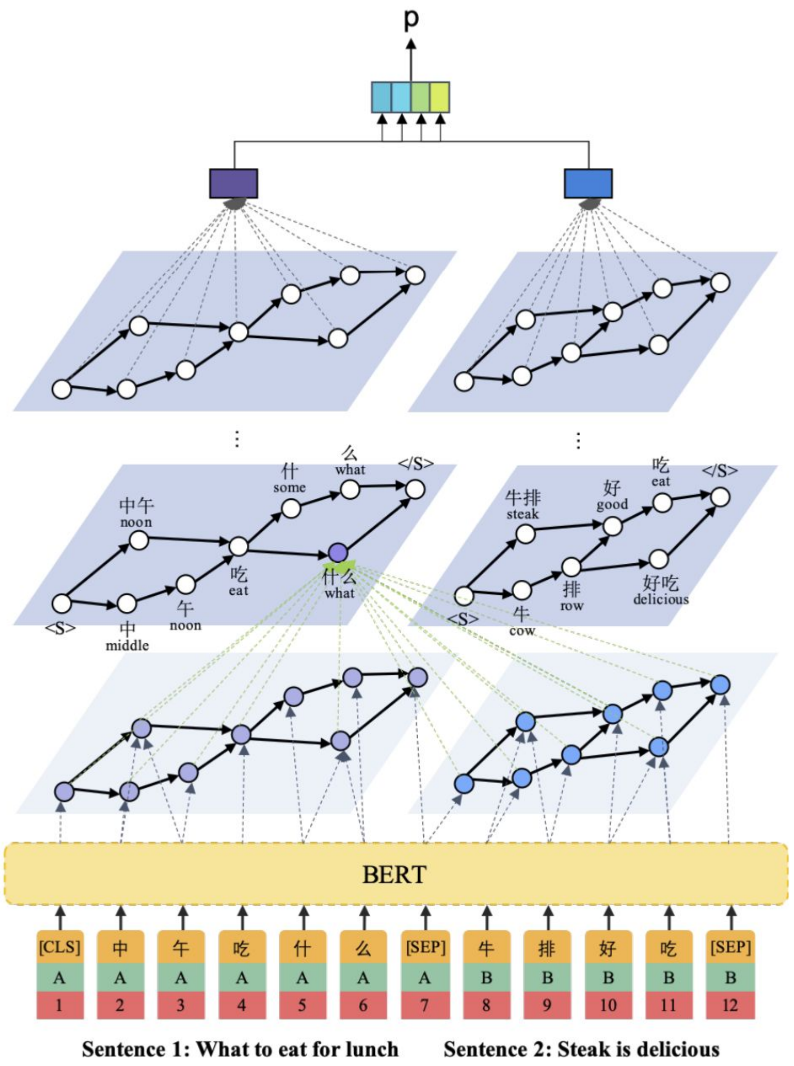
Lu Chen, Yanbin Zhao, Boer Lv, Lesheng Jin, Zhi Chen, Su Zhu, Kai Yu

ACL 2020

Motivation

- Chinese word segmentation is not as good as perfect
- This paper propose neural graph matching networks
- This framework is capable of dealing with multi-granular input information (word level and token level)
- Sentence matching: Matching the semantic of two sentences
-


Original Sentence：南京市长江大桥（ South | Capital | City | Long | River | Big | Bridge ）
Segment-1：南京市 / 长江大桥（Nanjing City / Yangtze River Bridge）
Segment-2：南京市 / 长江 / 大桥（ Nanjing City / Yangtze River / Bridge ）
Segment-3：南京 / 市长 / 江大桥（ Nanjing / Mayor / Daqiao Jiang）
Word Lattice
<S>
南 South
京 Capital
市 City
长 Long
江 River
大 Big
桥 Bridge
</S>
南京 Nanjing
市长 Mayor
江大桥 Daqiao Jiang
长江 Yangtze River
大桥 Bridge
南京市 Nanjing City
长江大桥 Yangtze River Bridge



Self hidden vectors

$$
\begin{aligned}
\mathbf{m}_i^{fw} &= \sum_{v_j \in \mathcal{N}_{fw}(v_i)} \alpha_{ij} \left(\mathbf{W}^{fw} \mathbf{h}_j^{l-1} \right), \\
\mathbf{m}_i^{bw} &= \sum_{v_k \in \mathcal{N}_{bw}(v_i)} \alpha_{ik} \left(\mathbf{W}^{bw} \mathbf{h}_k^{l-1} \right),
\end{aligned} \tag{1}
$$

Cross-sentence hidden vectors

$$
\begin{aligned}
\mathbf{m}_i^{b1} &= \sum_{v_m \in \mathcal{V}^b} \alpha_{im} \left(\mathbf{W}^{fw} \mathbf{h}_m^{l-1} \right), \\
\mathbf{m}_i^{b2} &= \sum_{v_q \in \mathcal{V}^b} \alpha_{iq} \left(\mathbf{W}^{bw} \mathbf{h}_q^{l-1} \right).
\end{aligned} \tag{2}
$$

Update representation

Multiperspective distances

$$d_k = \text{cosine}\left(\mathbf{w}_k^{cos} \odot \mathbf{m}_i^{self}, \mathbf{w}_k^{cos} \odot \mathbf{m}_i^{cross}\right),$$

Representation

$$\mathbf{h}_i^l = \text{FFN}\left(\left[\mathbf{m}_i^{self}, \mathbf{d}_i\right]\right),$$

Relation distribution

$$p = \text{FFN}\left(\left[\mathbf{g}^a, \mathbf{g}^b, \mathbf{g}^a \odot \mathbf{g}^b, |\mathbf{g}^a - \mathbf{g}^b|\right]\right)$$

Models	**BQ**		**LCQMC**	
	ACC.	**F1**	**ACC.**	**F1**
Text-CNN	68.5	69.2	72.8	75.7
BiLSTM	73.5	72.7	76.1	78.9
Lattice-CNN	78.2	78.3	82.1	82.4
BiMPM	81.9	81.7	83.3	84.9
ESIM-char	79.2	79.3	82.0	84.0
ESIM-word	81.9	81.9	82.6	84.5
GMN (Ours)	**84.2**	**84.1**	**84.6**	**86.0**
BERT	84.5	84.0	85.7	86.8
BERT-wwm	84.9	-	86.8	-
BERT-wwm-ext	84.8	-	86.6	-
ERNIE	84.6	-	87.0	-
GMN-BERT (Ours)	**85.6**	**85.5**	**87.3**	**88.0**