Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks

周郴莲

Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks

Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng and Jun Zhao
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
{yubo.chen,lhxu,kliu,djzeng,jzhao}@nlpr.ia.ac.cn

Abstract

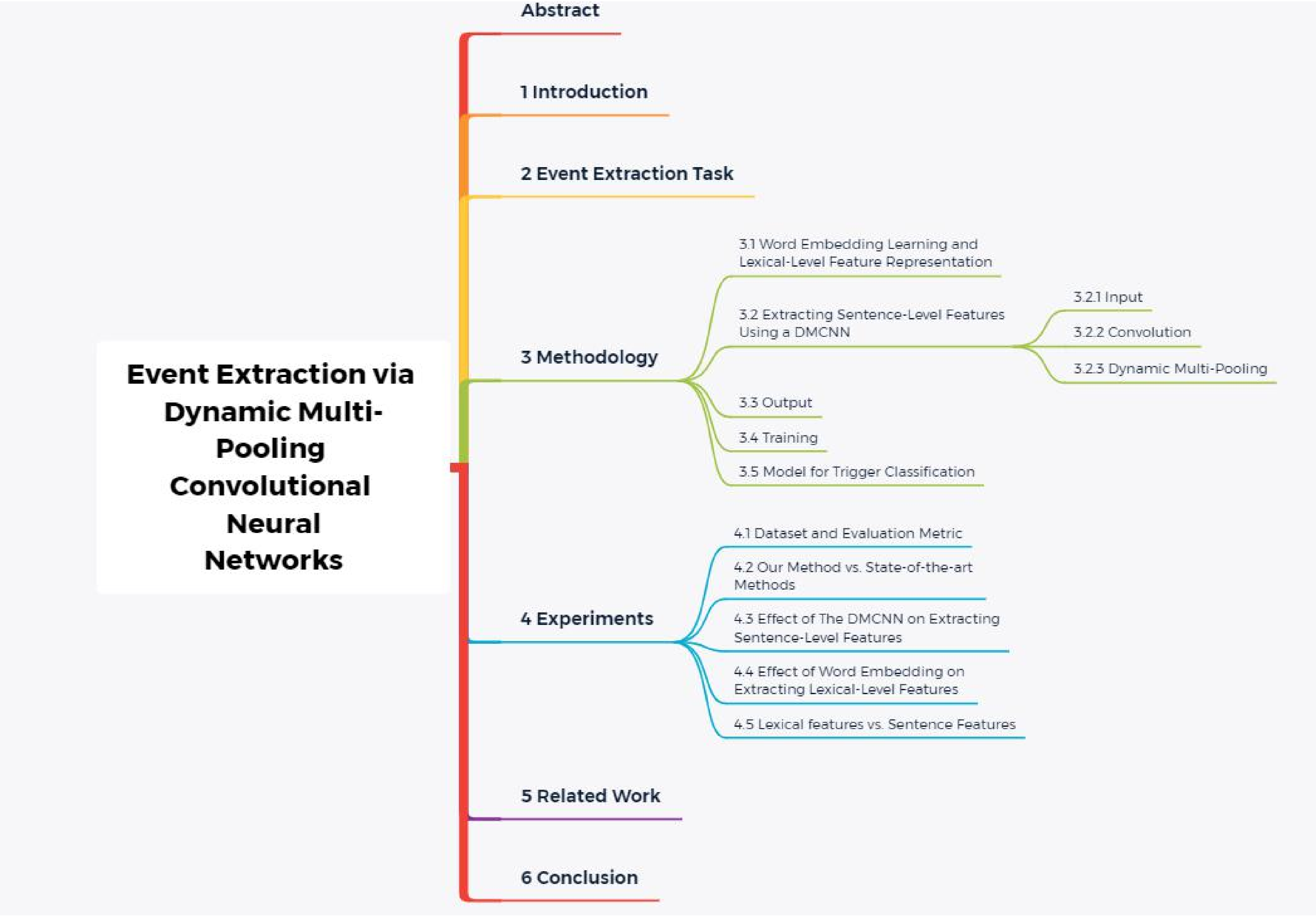
Traditional approaches to the task of ACE event extraction primarily rely on elaborately designed features and complicated natural language processing (NLP) tools. These traditional approaches lack generalization, take a large amount of human

knowledge. In general, we can divide the features into two categories: lexical features and contextual features.

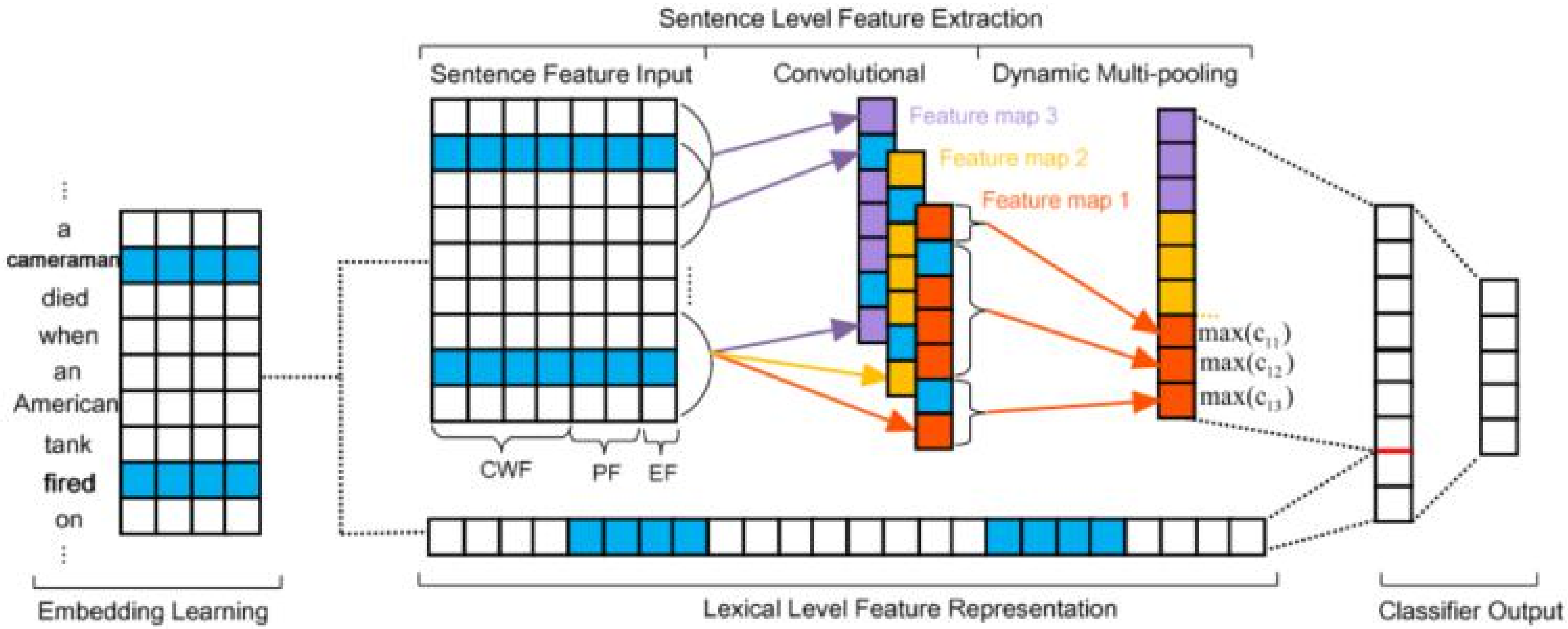
Lexical features contain part-of-speech tags (POS), entity information, and morphology features (e.g., token, lemma, etc.), which aim to capture semantics or the background knowledge of words. For example, consider the following sen-

S1: Obama **beats** McCain.
S2: Tyson **beats** his opponent .

Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks



Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks



$$x_{1:n} = x_1 \oplus x_2 \oplus \ldots \oplus x_n \qquad (3)$$

$$c_i = f(w \cdot x_{i:i+h-1} + b) \qquad (4)$$

$$c_{ji} = f(w_j \cdot x_{i:i+h-1} + b_j) \qquad (5)$$

$$p_{ji} = \max(c_{ji}) \qquad (6)$$

$$O = W_s F + b_s \qquad (7)$$

$$p(i|x, \theta) = \frac{e^{o_i}}{\sum_{k=1}^{n_1} e^{o_k}} \qquad (8)$$

$$J(\theta) = \sum_{i=1}^{T} \log p(y^{(i)}|x^{(i)}, \theta) \qquad (9)$$

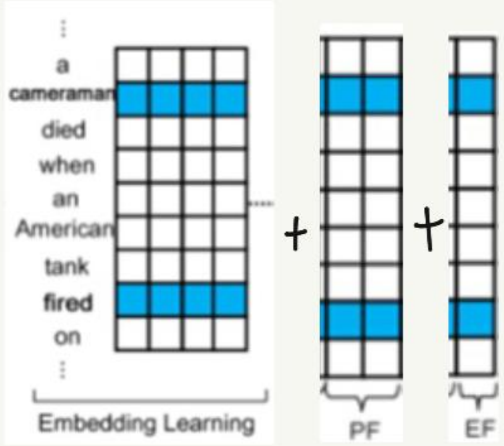
Target

In Baghdad, a cameraman died when an American tank fired on the Palestine Hotel

argument　　　　　　　　　　　　　　tigger

特征提取

4维



CWF　　PF　　EF

⇒ 句子所有单词的特征矢量 $x_1, x_2, \cdots, x_n$ ，d维

$x_i \in R^d$

h长度的window：$x_{i:i+h-1}$

$x_{i:i+h-1} \in R^{h \times d}$

window的句子：$x_{1:h}, x_{2:h+1}, \cdots, x_{n-h+1:n}$

⇒ 卷积 m个filter对每个window做卷积

$W = w_1, w_2, \cdots, w_m$

第j个filter，第i个window：$w_j \cdot x_{i:i+h-1} + b_j$

激活函数 f：$c_{ji} = f(w_j \cdot x_{i:i+h-1} + b_j)$

整个句子就有了 n-h+1 个输出，$c_{j1}, c_{j2}, \cdots, c_{jn-h+1}$

⇒ max-pooling 阶段

每一段进行 max-pooling：$p_{ji} = \max(c_{ji})$, $1 \le i \le 3$

3m个输出 $p \in R^{3m}$ 作为句子的特征

+ 词汇级特征 d_e 维 ⇒ $F = [L, P]$, $F \in R^{3m+d_e}$

线性分类器的输入

$O = W_s F + b_s$, $W_s \in R^{n_1 \times (3m+d_e)}$

argument role

⇒ 学习所有的参数

所有的 embedding 参数 E, PF_1, PF_2, EF

两个线性变换的参数 W, b, W_s, b_s

先算第i个 argument role：O_i → 非线性 exp 变换 → 除以所有的

$p(i|x,\theta) = \frac{e^{O_i}}{\sum_{k=1}^{n_1} e^{O_k}}$, $\theta = (E, PF_1, PF_2, EF, W, b, W_s, b_s)$

对于 $x^{(i)}$，正确的 argument role $y^{(i)}$ 的概率 $p(y^{(i)}|x^{(i)},\theta)$

最大 $\max p(y^{(i)}|x^{(i)},\theta)$ → log变化 $\max \log p(y^{(i)}|x^{(i)},\theta)$

都加起来 $\max\{J(\theta) = \sum_{i=1}^{T} \log p(y^{(i)}|x^{(i)},\theta)\}$ ⇒ 随机梯度下降

Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks

Methods	Trigger Identification(%)			Trigger Identification + Classification(%)			Argument Identification(%)			Argument Role(%)		
	P	R	F	P	R	F	P	R	F	P	R	F
Li's baseline	76.2	60.5	67.4	74.5	59.1	65.9	74.1	37.4	49.7	65.4	33.1	43.9
Liao's cross-event	N/A			68.7	68.9	68.8	50.9	49.7	50.3	45.1	44.1	44.6
Hong's cross-entity	N/A			72.9	64.3	68.3	53.4	52.9	53.1	51.6	45.5	48.3
Li's structure	76.9	65.0	70.4	73.7	62.3	67.5	69.8	47.9	56.8	64.7	44.4	52.7
DMCNN model	80.4	67.7	**73.5**	75.6	63.6	**69.1**	68.8	51.9	**59.1**	62.2	46.9	**53.5**

THANK YOU

感谢聆听