

# 基于区块提议网络的中文事件检测

Nugget Proposal Networks for Chinese Event Detection. In: ACL2018.

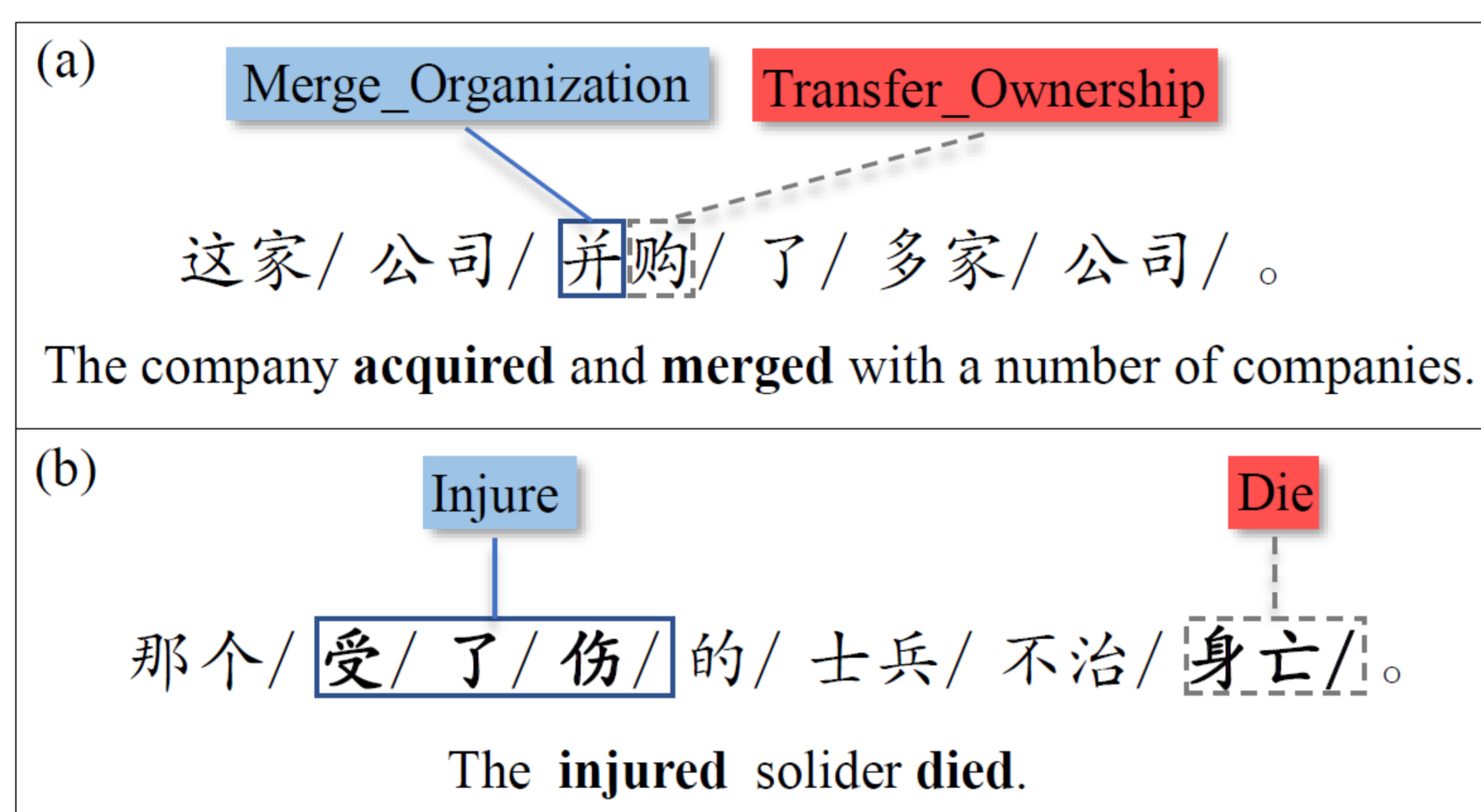
林鸿宇, 陆垚杰, 韩先培, 孙乐

联系人: 韩先培

联系方式: xianpei@iscas.ac.cn

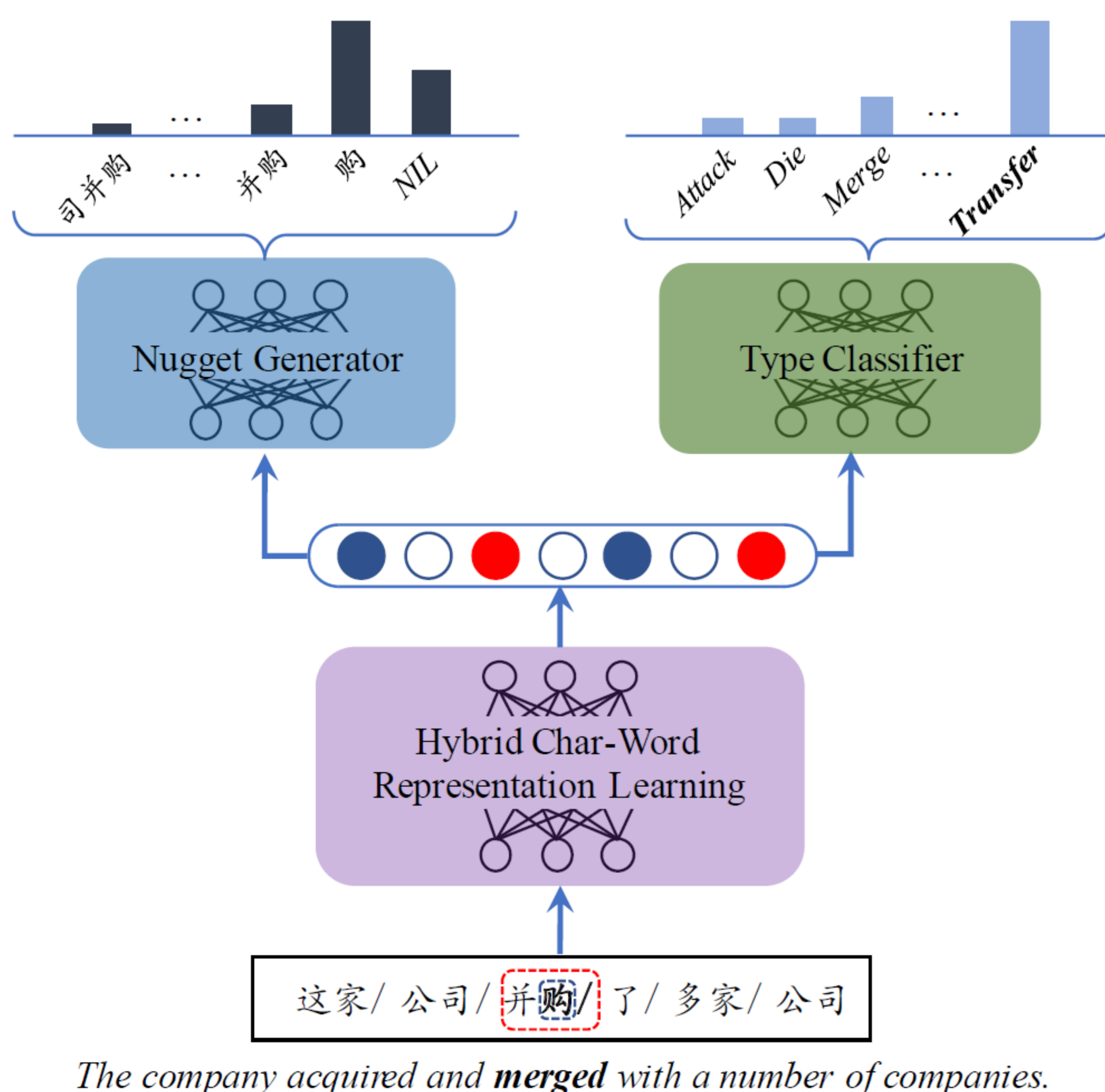
## 背景

- 深度学习方法在事件检测任务上已经取得了长足的进步
- 传统模型将事件检测视为词级别的分类问题
- 在中文等没有自然词分隔符的语言当中, 这些方法会受到**词与检测单元不匹配**问题的影响



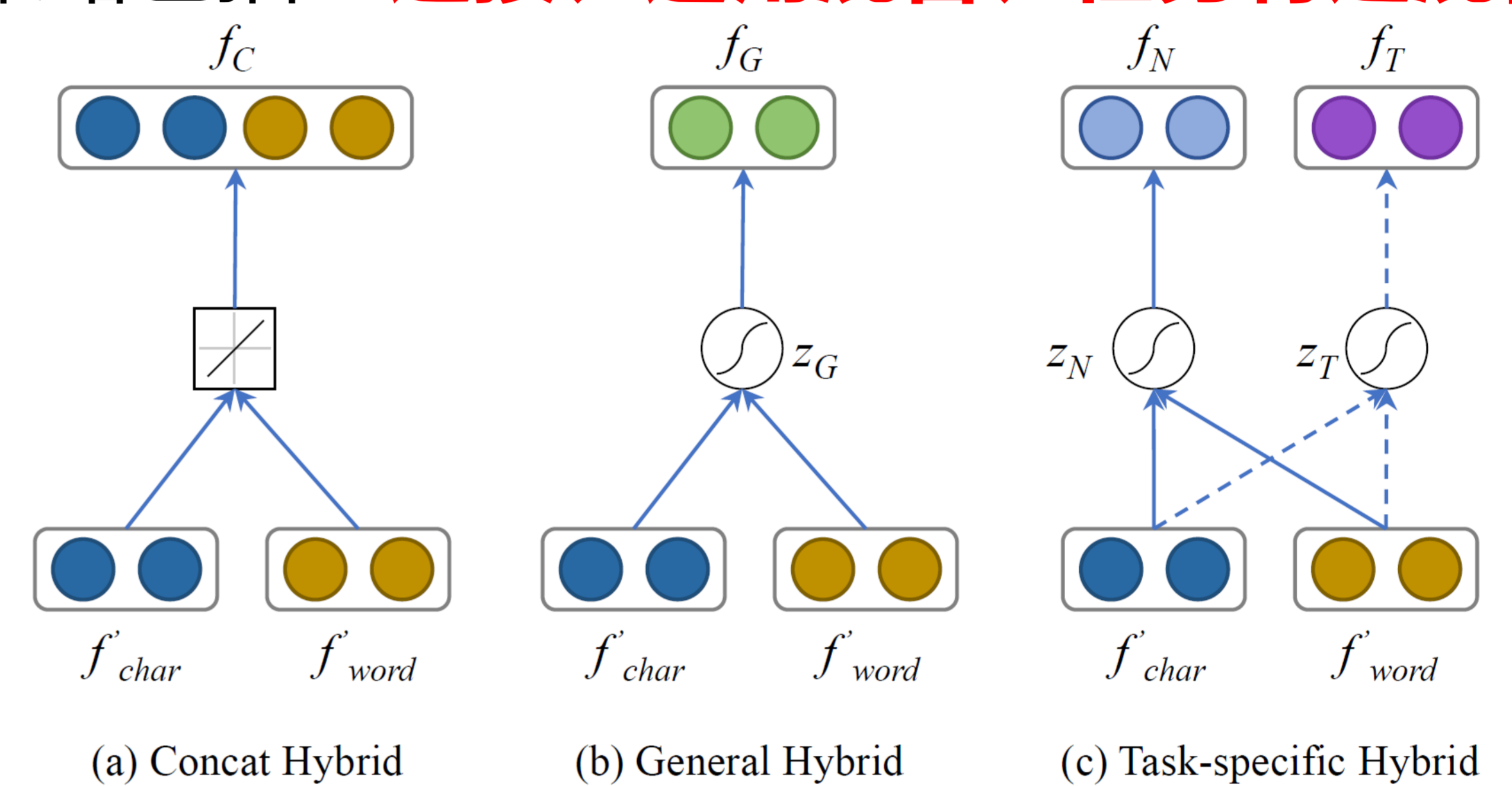
## 区块提议网络

- 建模中文触发词的内部**组合语义结构**
- 直接在每个字级别上识别整个触发词块**, 从而不受词边界的影响
- 使用**字与词混合的增强语义**表示来得到更充分的语义表征



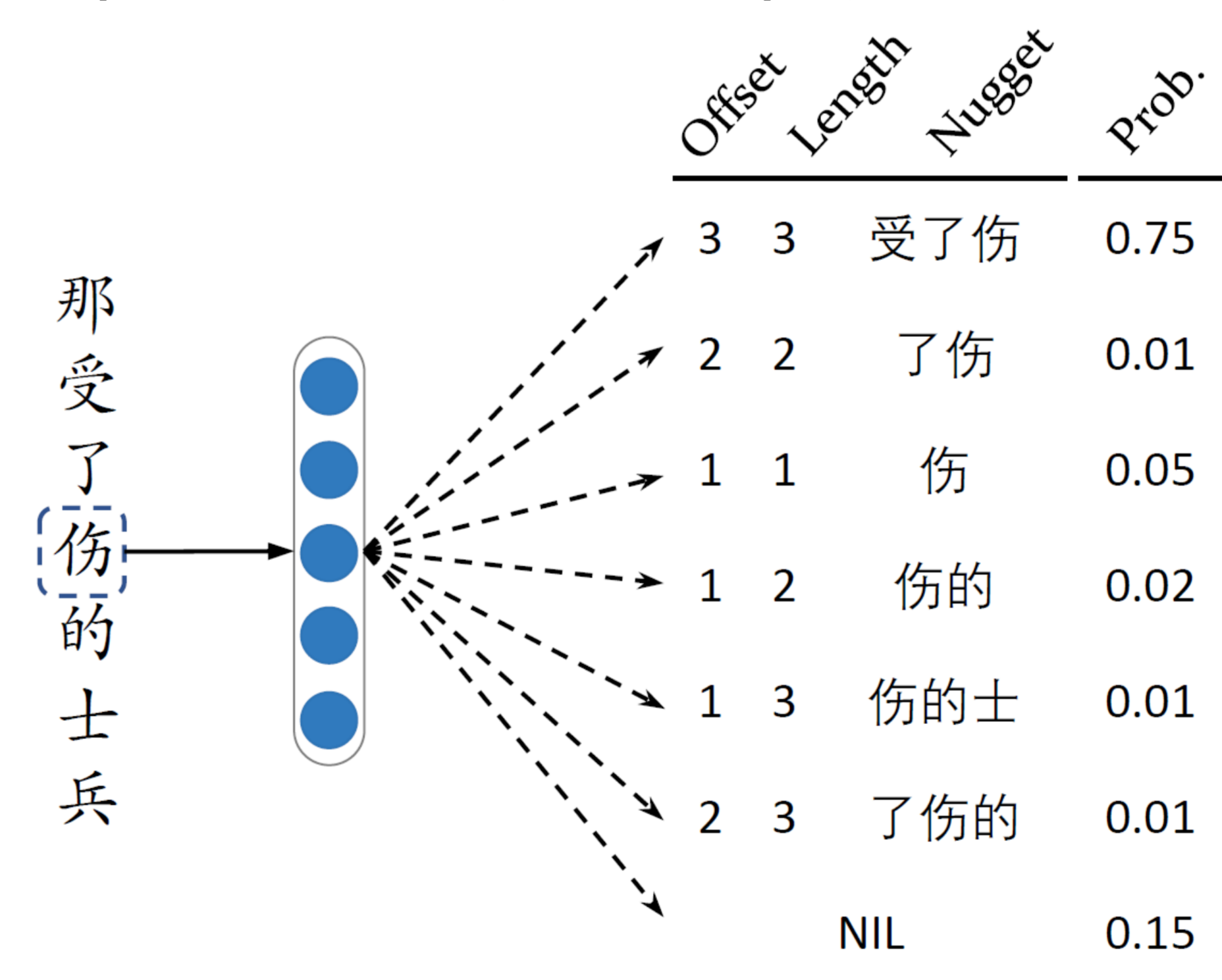
## 混合字-词表示学习

- 首先单独地学习字级别与词级别的表示
- 使用三种不同的混合策略得到最终的混合表示
- 策略包括: **连接、通用混合、任务特定混合**



## 触发词块生成器

- 在每一个字符上直接预测完整的出发词块, 避免词边界的影响
- 利用**冗余预测的策略**来处理不同字符上的结果冲突, 从而降低噪声, 提高模型的召回率



## 实验

- 在ACE2005和KBP2017两大事件检测数据集上得到了当前最好的结果

Model	ACE2005						KBPEval2017					
	Trigger Identification			Trigger Classification			Trigger Identification			Trigger Classification		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
FBRNN(Char)	61.3	45.6	52.3	57.5	42.8	49.1	57.97	36.92	45.11	51.71	32.94	40.24
DMCNN(Char)	60.1	61.6	60.9	57.1	58.5	57.8	53.67	49.92	51.73	50.03	46.53	48.22
C-BiLSTM*	65.6	66.7	66.1	60.0	60.9	60.4	-	-	-	-	-	-
FBRNN(Word)	64.1	63.7	63.9	59.9	59.6	59.7	65.10	46.86	54.50	60.05	43.22	50.27
DMCNN(Word)	66.6	63.6	65.1	61.6	58.8	60.2	60.43	51.64	55.69	54.81	46.84	50.51
HNN*	74.2	63.1	68.2	77.1	53.1	63.0	-	-	-	-	-	-
Rich-C*	62.2	71.9	66.7	58.9	68.1	63.2	-	-	-	-	-	-
KBP2017 Best*	-	-	-	-	-	-	<b>67.76</b>	45.92	54.74	<b>62.69</b>	42.48	50.64
NPN(Concat)	<b>76.5</b>	59.8	67.1	72.8	56.9	63.9	64.58	50.31	56.56	59.14	46.07	51.80
NPN(General)	71.5	63.2	67.1	67.3	59.6	63.2	63.67	51.32	56.83	57.78	46.58	51.57
NPN(Task-specific)	64.8	<b>73.8</b>	<b>69.0</b>	60.9	<b>69.3</b>	<b>64.8</b>	64.32	<b>53.16</b>	<b>58.21</b>	57.63	<b>47.63</b>	<b>52.15</b>