

# 基于字符词组特征与原型网络融合训练的事件分类

赵芝茵, 程良伦, 陈光明

广东工业大学计算机学院, 广东 广州  
Email: michel\_yyyle@163.com

收稿日期: 2021年3月20日; 录用日期: 2021年4月14日; 发布日期: 2021年4月22日

---

## 摘要

事件检测与分类任务, 包含两个步骤的子任务: 识别事件触发词和将其分类为正确的事件类型。在这项任务中首要关键的就是触发词的识别, 利用基于神经网络的模型来识别句子中的触发词是这些年的主流方法。然而, 当涉及到由语义结构不清和语义相近的字符和词组组成的句子时, 识别事件的触发词变得有些困难。本文提出一个融合字与词信息, 再通过原型网络来精确事件分类的模型: 输入融合字与词的信息的嵌入信息, 将各个组成的嵌入信息投影到一个高维的特征空间中, 对于每个维度类型的样本信息提取他们的均值作为聚类中心即原型, 使用欧几里得距离作为距离度量, 训练使得测试样本到自己类别原型的距离越近越好, 到其他类别原型的距离越远越好, 更精确地识别出句子所包含的触发词, 分辨出事件类型。

## 关键词

触发词检测, 事件分类, 原型网络, 神经网络

---

# Event Classification Based on Fusion Training of Character Phrase Features and Prototype Network

Zhiyin Zhao, Lianglun Cheng, Guangming Chen

School of Computer Science, Guangdong University of Technology, Guangzhou Guangdong  
Email: michel\_yyyle@163.com

Received: Mar. 20<sup>th</sup>, 2021; accepted: Apr. 14<sup>th</sup>, 2021; published: Apr. 22<sup>nd</sup>, 2021

## Abstract

The event detection and classification task consists of two-step subtasks: identifying the event trigger word and classifying it into the correct event type. The most important thing in this task is the recognition of trigger words. Using neural network-based models to identify trigger words in sentences is the mainstream method in these years. However, when it comes to sentences composed of characters and phrases with unclear semantic structure and similar semantics, it becomes difficult to identify the trigger words of the event. This paper proposes to train an n-dimensional prototype network that integrates the embedded information of the word information: input the embedded information of the fused word and word information, and project the embedded information of each composition into a high-dimensional feature space. For each dimension type, the sample information extracts their mean value as the cluster center or prototype, and uses the Euclidean distance as the distance metric. Training makes the test sample the closer to the prototype of its own category, the better, and the farther the distance to prototypes of other categories, the better. Accurately identify the trigger words contained in the sentence and distinguish the type of event.

## Keywords

Trigger Word Detection, Event Classification, Prototype Network, Neural Networks

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

事件检测(Event detection)的目标是检测文本中是否含有事件并对其进行分类。这项工作处理事件检测的任务,其目标是检测预定义事件的发生并对它们进行分类。例如,考虑下面这句话“在巴格达,美国坦克向巴勒斯坦饭店开枪时,一名摄影师丧生。”理想的事件检测系统应该识别两个事件,攻击事件和死亡事件(假设死亡和攻击都在预定义的事件集中)。

在事件检测阶段,触发字无法识别的问题会严重影响事件检测系统的性能。因为在中文中有一些词的含义是处于单字和单词中不一样的,触发器可能是一个单词或者一个单词的特定部分或包含多个单词。在这种情况下,基于单词的方法无法正确定位触发器,从而成为任务的严重限制。有一些基于特征的方法[1][2][3]被提出来缓解了这个问题,但是这些方法极大程度上依赖于人工标记。针对这一问题,相关人员提出了 Nugget Proposal Networks [4],该网络通过直接提出以每个字符为中心的整个触发器块来解释字触发器不匹配问题,而不用考虑字边界。然而,该网络的机制将候选触发器的范围限制在一个固定大小的窗口内,这是有些刻板的,并且存在触发器重复的问题。

该文章提出一种用于进行小样本分类任务的原型网络(Prototypical Network),其认为存在一种嵌入的表示方式,使得每一种类别对应的特征点都聚集在一个单独的原型表示附近。该文章为小样本和零样本设置建立了原型网络基于这一想法,作者采用了一种类似于聚类的方式,实现小样本分类任务。首先,作者采用一种简单的四级神经网络结构将输入的信息映射到高维的特征空间(度量空间),由于引入了 ReLU 激活函数因此该映射是非线性的。然后对于每一个类别,取其高维特征向量的均值作为原型

(Prototype, 可以理解为聚类中心)。最后计算测试样本对应的高维特征向量与各个原型之间的距离, 并利用 softmax 将其转化为概率值, 预测该样本的类别。但该文对一些更核心的问题没有给出解答, 比如输入部分采用简单的神经网络结构如何保证提取到的特征信息足够抽象等。

在这篇文章中, 我们提出将每个句子字符级的嵌入表示和词组级的嵌入表示结合到一起, 输入到原型网络中进行分类。将句子的字符级的嵌入表示和词组级的嵌入表示结合到一起, 能够更加抽象细腻表达每个字符元素的特征, 从而更准确识别出每个句子中决定事件类型的触发词。接着输入到原型网络中进行进一步的触发词识别和分类, 能弥补原型网络原有的训练上的不足, 抽象了原型网络的输入, 更深层次地识别出事件分类的触发词。

在基于小样本分类的原型网络中, 我们将字符信息与词组信息合并, 并分析原型网络模型中使用的距离计算函数。原型网络与聚类联系起来[5], 证明计算距离时用布雷格曼散度(如平方欧几里德距离), 使用类方法作为原型是合理的。根据以往经验可以确定的是, 距离的选择对聚类的准确度十分重要, 因为用欧几里德距离进行衡量标准比常用的余弦相似性更为有效和精确。在几项测试中, 我们实现了最先进的技能。将字层面与词层面的表示与原型网络小样本分类结合到一起, 改进了触发词识别与事件分类的效果。

## 2. 相关工作

### 2.1. 事件检测

事件检测任务要求检测注释数据中提到的特定类型的事件。之前工作中最常用的基准数据集是 ACE 2005 语料库。该语料库包括 8 种事件, 33 种子事件。我们忽略每一个事件和子事件之间的层次结构, 把这个数据集视为 33 种独立的事件类型。事件检测是自然语言处理的重要课题之一。已经为这项任务提出了许多方法。现有的 ACE 事件任务方法几乎都遵循监督范式。我们进一步将它们分为基于特征的方法和基于表示的方法。在基于特征的方法中, 已经使用了一组不同的策略来将分类线索转换成特征向量。Ahn [6]使用词汇特征(如全词)、句法特征(如依存特征)和外部知识特征(WordNet)来提取事件。受每个话语一个意义的假设的启发, [7]等学者将来自相关文献的全局证据与用于事件提取的局部决策相结合。为了从文本中获取更多的线索, [8] [9] [10]等学者提出了 ACE 事件任务的跨事件和跨实体推理。[11]等学者提出了一个联合模型来捕捉触发器和参数的组合特征。[12]等学者提出了一种全局推理方法, 利用潜在的局部和全局信息进行事件检测。近年来, 基于表征的方法主导了研究。在这个范例中, 候选事件提及由嵌入表示, 嵌入通常被馈送到神经网络中。[13]和[14]是这一范式的第一个工作。他们的模型基于卷积神经网络。为了模拟触发器和参数的依赖性, [15]等学者提出了一种基于递归神经网络的联合事件提取方法。[16]等学者提出通过监督注意机制对事件检测中的论点信息进行编码。

### 2.2. 近邻成分分析算法(NCA 算法)

距离测度学习算法在机器学习中非常重要, 此算法目的在于学习一种距离测度  $d(x_i, x_j)$  来表征  $x_i$  与  $x_j$  之间的某种相似度。其本质是通过对本样本进行线性或非线性变换获取另一种更有类别区分度的表示形式, 更好地反映了样本的空间特性, 在变换过程中保持了样本空间的流形。NCA 算法就是一种简单有效的距离测度学习算法。该算法随机选择近邻, 通过优化留一法(Leave-one-out, LOO)的交叉检验结果来求得马氏距离中的变换矩阵。

## 3. 本文方法

本文将触发词识别与事件分类视为一项序列标记任务。对于每个字符, 模型应该识别它是否是一个

触发器的一部分，并将该触发器正确地分类为一个特定的事件类型。我们的模型架构主要包括以下三个部分：

- (1) 分层表示学习，它揭示字符级、单词级和意义级嵌入向量。
- (2) 触发感知特征提取器，通过树形结构 LSTM 模型自动提取不同层次的语义特征。
- (3) 将结合不同层次的特征非线性映射到嵌入空间，计算各特征集群之间的距离，求出均值作为类的原型，然后在测试集中对嵌入的查询点进行分类。

### 3.1. 分层表示学习

给定输入序列  $S = \{c_1, c_2, \dots, c_N\}$ ，其中  $c_i$  表示序列中的第  $i$  个字符。使用 Skip-Gram 模型 SkipGram 方法[17]，在字符级，每个字符将被表示为  $x^c$  向量：

$$x_i^c = e(c_i) \quad (1)$$

在单词级，输入序列  $S$  也是  $S = \{w_1, w_2, \dots, w_M\}$ ，其中单个词  $w_i$  表示序列中的第  $i$  个词语。在本文中，我们将使用  $b$  和  $e$  两个索引来表示一个单词的开头和结尾。在单词级中表示为：

$$x_{b,e}^w = e(w_{b,e}) \quad (2)$$

### 3.2. 触发感知特征提取器

触发感知特征提取器是我们模型的核心组件。经过训练，提取器的输出是输入句子的隐藏状态向量  $h$ 。传统 LSTM 是递归神经网络(RNN)的扩展，在 RNN 上增加了门(gates)来控制信息。按传统的来说，LSTM 有以下基本门：输入门  $i$ 、输出门  $o$  和遗忘门  $f$ 。它们共同控制哪些信息要被保留、遗忘和输出。所有三个门都配有相应的权重矩阵  $W$ 。当前窗口状态  $c$  记录了所有流到当前时间的历史信息。因此，基于字符级别的 LSTM 函数是：

$$\begin{bmatrix} i_i^c \\ o_i^c \\ f_i^c \\ \tilde{c}_i^c \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} \left( W^{cT} \begin{bmatrix} x_i^c \\ h_{i-1}^c \end{bmatrix} + b^c \right) \quad (3)$$

$$c_i^c = f_i^c \odot c_{i-1}^c + i_i^c \odot \tilde{c}_i^c \quad (4)$$

$$h_i^c = o_i^c \odot \tanh(c_i^c) \quad (5)$$

其中  $h_i^c$  是隐藏状态向量。

本文使用的触发词感知网格 LSTM 是 LSTM 和 lattice LSTM 的扩展。假定字符和单词有  $K$  个含义，第  $i$  个字符  $c_i$  的第  $j$  个含义的表示为  $s_j^{c_i}$ 。

#### 1) 整合字符的多义信息

使用附加的 LSTMCell 整合该字符的所有含义，因此多义字符  $c_i$  的 cellgate 计算是：

$$\begin{bmatrix} i_j^{c_i} \\ f_j^{c_i} \\ \tilde{c}_j^{c_i} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \tanh \end{bmatrix} \left( W^{cT} \begin{bmatrix} s_j^{c_i} \\ h^{c_{i-1}} \end{bmatrix} + b^c \right) \quad (6)$$

$$c_j^{c_i} = f_j^{c_i} \odot c^{c_{i-1}} + i_j^{c_i} \odot \tilde{c}_j^{c_i} \quad (7)$$

其中  $c_j^{c_i}$  是第  $i$  个字符的第  $j$  个含义的 cell 状态， $c^{c_{i-1}}$  是第  $i-1$  个字符的最终 cell 状态。为了获得字符的

cell 状态，使用了一个附加的字符含义门：

$$g_j^{c_i} = \sigma \left( W^T \begin{bmatrix} x_i^c \\ c_j^{c_i} \end{bmatrix} + b \right) \quad (8)$$

然后所有的含义需要被动态整合到一个临时的 cell 状态，如下所示。其中  $\alpha_j^{c_i}$  是归一化后的字符含义门：

$$c^{*c_i} = \sum_j^K \alpha_j^{c_i} \odot c_j^{c_i} \quad (9)$$

$$\alpha_j^{c_i} = \frac{\exp(g_j^{c_i})}{\sum_k^K \exp(g_k^{c_i})} \quad (10)$$

## 2) 整合词的多义信息

式(9)通过合并字符所有的含义信息，得到了临时的 cell 状态  $c^{*c_i}$ ，但是，还需要考虑词语级(word level)的信息， $s_j^{w_{b,e}}$  表示词  $w_{b,e}$  的第  $j$  个含义的表示。和字符类似，使用 LSTMCell 计算每个单词的 cell state:

$$\begin{bmatrix} i_j^{w_{b,c}} \\ f_j^{w_{b,c}} \\ c_j^{w_{b,c}} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \tanh \end{bmatrix} \left( W^T \begin{bmatrix} s_j^{w_{b,c}} \\ h^{c_{b-1}} \end{bmatrix} + b^c \right) \quad (11)$$

$$c_j^{w_{b,c}} = f_j^{w_{b,c}} \odot c^{c_{b-1}} + i_j^{w_{b,c}} \odot \tilde{c}_j^{w_{b,c}} \quad (12)$$

和式(8)~(10)类似，整合所有 senses 的 cells 信息，得到词的 cell state:

$$g_j^{w_{b,c}} = \sigma \left( W^T \begin{bmatrix} x_b^c \\ c_j^{w_{b,e}} \end{bmatrix} + b \right) \quad (13)$$

$$c^{w_{b,e}} = \sum_j^K \alpha_j^{w_{b,e}} \odot c_j^{w_{b,e}} \quad (14)$$

$$\alpha_j^{w_{b,e}} = \frac{\exp(g_j^{w_{b,e}})}{\sum_k^K \exp(g_k^{w_{b,e}})} \quad (15)$$

## 3) 合并字符信息和词信息

对于字符  $c_i$ ，临时的 cell 状态  $c^{*c_i}$  包含了所有 sense 的信息。通过式(14)也可以计算出所有以索引  $i$  结尾的词的 cell 状态，记为  $\{c^{w_{b,i}} \mid b \in [1, i], w_{b,i} \in D\}$  ( $D$  表示词典)。为了保证对应的信息可以流到  $c_i$  最终的 cell 状态，使用一个额外的门  $g_{b,i}^m$  整合字符 cells 和词 cells:

$$g_{b,i}^m = \sigma \left( W^\Pi \begin{bmatrix} x_i^c \\ c_{b,i}^w \end{bmatrix} + b^l \right) \quad (16)$$

字符  $c_i$  最终的 cell 状态计算如下，其中  $\alpha^{w_{b,i}}$  和  $\alpha^{c_i}$  分别是 word gate 和 character gate 归一化后的值：

$$c^{c_i} = \sum_{b \in \{b \mid w_{b,i}^d \in D\}} \alpha^{w_{b,i}} \odot c^{w_{b,i}} + \alpha^{c_i} \odot c^{*c_i} \quad (17)$$

因此，最终的  $c^{c_i}$  可以动态地表示多义的字符和词。如式(5)所示，得到的  $c^{c_i}$  再过一个输出门，得到输出，也就是原型网络的输入。

### 3.3. 原型网络分类

该网络由四个卷积块构成, 每个卷积块包含一个 64 通道的  $3 * 3$  的卷积层, 一个批规范化层, 一个 ReLU 激活层和一个  $2 * 2$  的最大池化层。

距离度量方式本文选用的是 Bregman 散度中的平方欧氏距离, 这是因为 Bregman 散度具备一种优良的性质, 对于一系列特征点, 特征空间中与所有点之间的平均距离最近的点是该系列特征点的均值。正是由于该原因, 我们可以直接利用每个类别特征向量的均值作为该类别对应的原型。采用平方欧氏距离作为距离度量方式其效果要优于余弦距离, 我们认为利用 Bregman 散度作为距离度量方式可以将小样本的分类问题转化为一种线性分类模型, 而必要的非线性信息包含在嵌入的神经网络中了。

对于训练样本与原型之间的距离  $d$  取负数, 并利用 softmax 函数转化为概率值  $p$ , 对概率值取负对数作为损失。目标是使样本与正确类别原型之间的距离越来越近, 与其他类别原型之间的距离越来越远。

本文采用一种称为 Episode 的小批量梯度下降训练方法, 对于每一个批次随机选择  $N_c$  个类别的样本, 对于每个类别选择  $N_s$  个样本作为支持训练集(Support Set), 用于计算原型。对于每个类别, 除去作为支持训练集的样本, 剩余的样本中选取  $N_o$  个样本作为测试点, 用于计算与原型之间的距离, 并得到损失, 更新网络的参数。作者通过实验发现, 用于训练的样本类别数量应该多于测试样本的类别数量, 而每个类别训练样本的数量应该与测试样本的数量相同, 这样能够改善小样本的分类效果。原作者认为是由于训练中采用更多类别的样本使得训练任务更加困难, 减弱了过拟合问题, 提高了模型的泛化能力。

## 4. 实验

### 4.1. 数据集与实验设置

#### 4.1.1. 数据集

本文在两个真实数据集上进行了一系列实验: ACE2005 中文数据集(ACE 2005)和 TAC KBP 2017 事件金块检测评估数据集(KBP 2017)。为了更好的比较, 我们使用了与以前作品相同的数据分割[18] [19] [20]。具体来说, ACE2005 数据集包含 697 篇文章, 其中 569 篇用于培训, 64 篇用于验证, 其余 64 篇用于测试。对于 KBP2017 中文数据集(LDC2017E55), 我们遵循与[19]等学者相同的设置, 分别使用 506/20/167 文档作为训练/开发/测试集。

#### 4.1.2. 评估标准

评估标准选择标准微平均精度(Standard micro-averaged Precision)、召回率(Recall)和 F1 分数。对于 ACE2005, 计算方法与[1]相同。

#### 4.1.3. 超参数设置

我们通过在验证数据集上进行网格搜索来调整模型的参数。Adam [18]利用学习率衰减作为优化器。文字和感官的嵌入大小都是 50。为了避免过拟合, 系统中使用了 rate 机制[13], dropout rate 设置为 0.5。我们通过提前停止使用验证数据集上的 F1 结果来选择最佳模型。由于影响有限, 我们遵循其他超参数的经验设置。

## 4.2. 结果

在这一部分, 我们将我们的模型与以前最先进的方法进行比较。之前提出过的最先进的模型有以下几种。

DMCNN: 提出了一种动态多池卷积神经网络, 它根据事件触发器和参数使用动态多池层来保留更多关键信息。

**HBTNGMA**: 提出了一个具有门控多级注意机制的层次化和偏向性标注网络, 以集中整合句子级和文档级信息。

**NPN**: 提出了一种通过自动学习触发器内部组成结构来解决触发器不匹配问题的综合模型。

从表 1 结果中, 我们可以观察到:

(1) 对于 ACE2005 和 KBP2017, 我们的方法显著优于其他提出的模型, 在两个数据集上取得了最好的结果。这表明添加两种级别的表示信息可以更加精确触发词的识别。此外, 由于加入了多重级别的信息, 高精度识别出触发词, 此模型能够更加精确识别出事件类型。

(2) 通过原型网络的聚类将所有候选单词的快捷路径与当前字符联系起来, 该模型可以有效地利用字符和单词信息, 从而缓解触发词不匹配的问题。此外, 使用欧几里德距离大大提高了余弦距离的性能。这种效应对于原型网络更为明显, 在原型网络中, 因为余弦距离与布雷格曼散度无关, 将类原型计算为表示支持点的平均值更自然地适用于欧几里德距离。

**Table 1.** On the data set ACE2005 and the data set KBP2017, the training results of DMCNN and HBTNGMA based on the character level, the training results of DMCNN and HBTNGMA based on the word level, and the NPN of the joint training of words and words are compared with the results of this method

**表 1.** 在数据集 ACE2005 上与数据集 KBP2017 上, 基于字符级提取的 DMCNN 与 HBTNGMA 的训练结果, 基于词语级的 DMCNN 与 HBTNGMA 的训练结果, 以及字词混合联合训练学习的 NPN 与本文方法的结果对比

模型		ACE2005			KBP2017		
		P	R	F1	P	R	F1
字符级	DMCNN	60.10	61.60	60.90	53.67	49.92	51.73
	HBTNGMA	41.67	59.29	48.94	40.52	46.76	43.41
词语级	DMCNN	66.60	63.60	65.10	60.43	51.64	55.69
	HBTNGMA	54.29	62.82	58.25	46.92	53.57	50.02
字词	NPN	70.63	64.74	67.56	58.03	59.91	58.96
混合表征	我们的方法	68.12	75.46	71.60	66.71	59.85	63.09

## 5. 结论

我们提出了字符词组特征与原型网络融合训练事件分类的模型来解决提高触发词识别精确度和明确事件类型分类的问题。字符级与词组级的表示学习可以同时解决触发词不匹配和多义词触发的问题, 通过分层表示学习和触发感知特征提取器, 再加入原型网络改进事件分类, 能够有效地利用多层次信息并学习深层语义特征。在两个真实数据集上的实验表明, 该模型能够有效地解决这两个问题, 并产生比各种神经网络模型更好的效果。在未来的工作中, 我们将在更多语种上进行实验, 努力实验出更好的效果。

## 致 谢

在本论文的写作过程中, 我的导师程良伦教师倾注了很多的心血, 从选题到做实验, 从写作提纲, 到一遍又一遍地指出每稿中的具体问题, 严格把关, 循循善诱, 在此我表示衷心感激。同时我还要感激在我学习期间给我极大关心和支持的各位教师以及关心我的同学和朋友。

## 参考文献

- [1] Chen, Z. and Ji, H. (2009) Language Specific Issue and Feature Exploration in Chinese Event Extraction. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for*

- Computational Linguistics, Companion Volume: Short Papers*, Boulder, June 2009, 209-212.  
<https://doi.org/10.3115/1620853.1620910>
- [2] Chen, Y.B., Xu, L.H., Liu, K., Zeng, D.J. and Zhao, J. (2015) Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, July 2015, 167-176. <https://doi.org/10.3115/v1/P15-1017>
- [3] Zeng, Y., Yang, H.H., Feng, Y.S., Wang, Z. and Zhao, D.Y. (2016) A Convolution BiLSTM Neural Network Model for Chinese Event Extraction. *Natural Language Understanding and Intelligent Applications: 5th CCF Conference on Natural Language Processing and Chinese Computing, NLPCC 2016, and 24th International Conference on Computer Processing of Oriental Languages, ICCPOL 2016*, Kunming, 2-6 December 2016, 275-287.  
[https://doi.org/10.1007/978-3-319-50496-4\\_23](https://doi.org/10.1007/978-3-319-50496-4_23)
- [4] Lin, H.Y., Lu, Y.J., Han, X.P. and Sun, L. (2018) Nugget Proposal Networks for Chinese Event Detection. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, July 2018, 1565-1574.
- [5] Banerjee, A., Merugu, S., Dhillon, I.S. and Ghosh, J. (2005) Clustering with Bregman Divergences. *Journal of Machine Learning Research*, **6**, 1705-1749.
- [6] Ahn, D. (2006) The Stages of Event Extraction. *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, Sydney, July 2006, 1-8. <https://doi.org/10.3115/1629235.1629236>
- [7] Ji, H. and Grishman, R. (2008) Refining Event Extraction through Cross-Document Inference. *Proceedings of ACL-08: HLT*, Columbus, June 2008, 254-262.
- [8] Gupta, P. and Ji, H. (2009) Predicting Unknown Time Arguments Based on Cross-Event Propagation. *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, Suntec, August 2009, 369-372.  
<https://doi.org/10.3115/1667583.1667697>
- [9] Liao, S.S. and Grishman, R. (2010) Using Document Level Cross-Event Inference to Improve Event Extraction. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, 11-16 July 2010, 789-797.
- [10] Hong, Y., Zhang, J.F., Ma, B., Yao, J.M., Zhou, G.D. and Zhu, Q.M. (2011) Using Cross-Entity Inference to Improve Event Extraction. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Stroudsburg, 1127-1136.
- [11] Li, Q., Ji, H. and Huang, L. (2013) Joint Event Extraction via Structured Prediction with Global Features. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers*, 73-82.
- [12] Liu, S.L., Liu, K., He, S.Z. and Zhao, J. (2016) A Probabilistic Soft Logic Based Approach to Exploiting Latent and Global Information in Event Classification. *The Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, 12-17 February 2016.
- [13] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, **15**, 1929-1958.
- [14] Nguyen, H.T. and Grishman, R. (2015) Event Detection and Domain Adaptation with Convolutional Neural Networks. *The 53rd Annual Meeting of the Association for Computational Linguistics*, Beijing, 26-31 July 2015.  
<https://doi.org/10.3115/v1/P15-2060>
- [15] Nguyen, H.T. and Grishman, R. (2016) Modeling Skip-Grams for Event Detection with Convolutional Neural Networks. *The 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, 1-5 November 2016.  
<https://doi.org/10.18653/v1/D16-1085>
- [16] Liu, S.L., Chen, Y.B., Liu, K. and Zhao, J. (2017) Exploiting Argument Information to Improve Event Detection via Supervised Attention Mechanisms. *The 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, 30 July-4 August 2017.
- [17] Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013) Efficient Estimation of Word Representations in Vector Space.
- [18] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
- [19] Lin, H.Y., Lu, Y.J., Han, X.P. and Sun, L. (2018) Nugget Proposal Networks for Chinese Event Detection. arXiv preprint arXiv:1805.00249.
- [20] Feng, X.C., Qin, B. and Liu, T. (2018) A Language-Independent Neural Network for Event Detection. *Science China Information Sciences*, **61**, Article ID: 092106. <https://doi.org/10.1007/s11432-017-9359-x>