

通过 Δ 表示学习对事件识别进行蒸馏区分和泛化知识

陆耀杰^{1,3}, 林宏宇^{1,3}, 韩先培^{1,2*}, 孙乐^{1,2}

¹ 中国信息处理实验室 ² 中国科学院计算机科学技术研究所软件国家重点实验室, 北京

³ 中国科学院大学, 北京

{ YajeIE2017, HuYu2016, ChanPipe, Sunle } @ ISCAS.AC.CN

摘要

事件检测系统依靠判别知识来区分模棱两可的触发词, 并依靠泛化知识来检测看不见/稀疏的触发词。当前的神经事件检测方法着重于以触发为中心的表示形式, 该方法在提炼歧视知识方面表现良好, 但在学习泛化知识方面表现不佳。为了解决这个问题, 本文提出了一种 Δ 学习方法, 通过有效地去耦, 增量学习和自适应融合事件表示来消除歧视和泛化知识。实验表明, 对于看不见/稀疏的触发词, 我们的方法明显优于以前的方法, 并且在 ACE2005 和 KBP2017 数据集上均达到了最新的性能。

1 介绍

事件检测 (ED) 旨在识别特定事件类型的触发器。例如, ED 系统将在“在巴勒斯坦酒店上开枪的美国坦克”一句中识别出被触发为攻击事件触发事件。事件检测在自动内容提取中起着重要的作用 (安, 2006), 信息检索 (艾伦, 2012) 和文字理解 (钱伯斯和尤拉夫斯基, 2008)。

由于自然语言表达的歧义性和多样性 (Li 等., 2013; 阮和 格里什曼, 2015), 那么一种有效的方法应能够提炼出用于事件检测的判别知识和概括知识。歧视知识旨在区分不同情况下的歧义触发因素。如图所示 1, 为了将 S4 中触发的触发识别为 EndPosition 触发而不是 Attack 触发, ED 系统需要从 S1 和 S2 中提取通常同时发生 (触发, Attack) 的识别知识

——通讯作者——

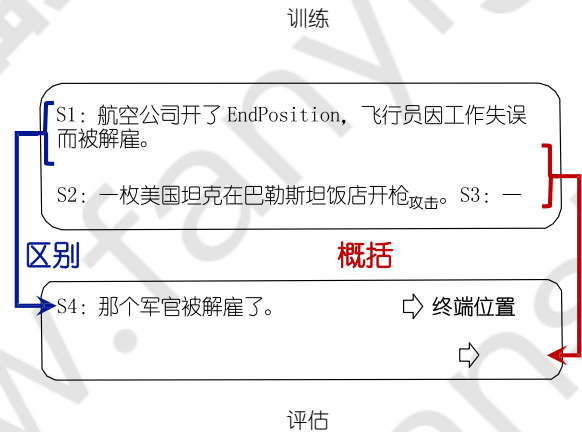


图 1: 事件实例的例子。识别出歧义的单词需要识别知识, 而识别看不见的单词被入侵则需要通用知识。

与 {坦克, 死亡, 敌人, ...} 和 (被解雇, EndPosition) 通常与 {工作, 过失, 工作, ...}. 与歧视知识不同, 概括知识旨在检测看不见的或稀疏的标记的触发器, 因此需要在不同的触发器字之间传输。例如, 为了将 S5 中未看见的单词识别为 Attack 触发器, ED 系统需要从 S3 中提取出广义的 Attack 模式 “[Trigger] to death”。

目前, 大多数神经网络 ED 方法 (Chen 等., 2015; 阮和格里什曼, 2015, 2016; 段等., 2017; 杨和米切尔, 2017) 在提取歧视知识方面做得很好, 但在提取概括性知识方面做得不好。表 1 显示了稀疏 (00V / 00L) 和密集 (Other) 标记触发字的几种模型的性能。这些模型在密集标记的触发词上很好地工作, 即它们具有良好的辨别能力。但是它们在看不见/稀疏标记的触发词上表现不佳, 即, 泛化能力很差。这是因为这些方法大多以触发器为中心, 因此很难很好地推广到稀疏/看不见的单词。此外, 缺乏

计算语言学协会第 57 届年会论文集, 第 4366-4376 页, 意大利佛罗伦萨, 2019 年 7 月 28 日至 8 月 2 日. 2019 年 Qc 计算语言学协会

模型	奥夫	奥奥尔	其他
德姆 cnn	34.3	8.8	76.1
双 LSTM	35.3	9.3	75.5
ELMo	31.3	9.0	75.7

表 1: 先前针对不同类型的触发器的方法的 F1 分数 (ACE2005), 其中 OOV 词是训练语料库中的词汇外词, OOL 词是标签外的词, 即, 其实例为 (单词, 事件类型) 从未出现在训练语料库中, 但单词不是 OOV. DMCNNChen 等., 2015) 是指基于动态多池的 CNN; Bi LSTM 段等., 2017) 是指基于双向 LSTM 的 RNN. ELMo 指由以下人员提出的与任务无关的固定词表示形式彼得斯等. (2018).

大规模的训练数据也限制了学习模型的泛化能力。表 1 还显示了使用通用的预训练单词表示法- ELMo (彼得斯等., 2018). 我们可以看到, 这种与任务无关的以词法为中心的表示与针对特定任务的表示几乎实现了相同的性能。

在本文中, 我们提出了一种 Δ 表示学习方法, 该方法可以逐步提炼判别和泛化知识以进行事件检测。 Δ 表示学习的目的是为事件表示解耦, 学习和融合可变的 Δ -部分, 而不是学习单个全面的表示。具体来说, 我们将事件表示 $r_{\text{预计起飞时间}}$ 解耦 电子束 分为三部分 $r_{\text{预计起飞时间}}$ r_w $r_{\text{EB}} r_g$ (部分 2),

其中 r_w 是触发词的预训练词表示, r_d 是特定于词法的事件表示, 它捕获用于区分歧义触发器的辨别知识, r_g 是无词法的事件表示, 其捕获用于检测非正常事件的泛化知识。可见/稀疏触发器, 电子束 是融合函数融合不同的零件。这里的 r_d 和 r_g 是我们表示的 Δ -部分, 即, 它们是从 r_w 开始独立学习的, 旨在捕获增量知识以进行事件检测。为了逐步学习 Δ -部分 r_d 和 r_g , 我们提出了 Δ -学习框架 (本节 3), 即设计词汇增强型 Δ 学习算法以学习与事件相关和词汇相关部分的区分知识 r_d , 而设计词汇对抗性 Δ 学习以学习泛化知识 r_g 这是事件相关但与词汇无关的部分。最后是

Δ-学习

解耦表示

触发候选
选人

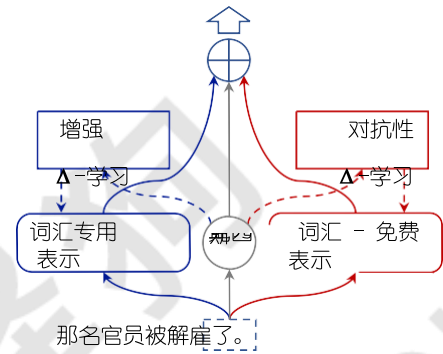


图 2: 我们的Δ学习方法的框架。虚线表示学习过程。实线表示事件检测过程。

重刑 2.3 提出了将这些学习的表示自适应融合的方法。数字 2 显示了我们方法的体系结构。

我们进行实验¹ 在两个标准上
事件检测数据集: ACE2005² 和 TAC KBP 2017
事件块检测检测评估³ (KBP2017)。实验结果表明, 该方法可以显著提高稀疏标记触发条件下的性能, 并在密集标记触发条件下保持较高的性能。

本文的主要贡献是:

1. 我们提出了一个新的表征学习框架-Δ-learning, 它可以在表征学习期间逐步提炼歧视知识和泛化知识。由于自然语言表达中的歧义和多样性问题在 NLP 中很常见, 因此我们的框架可以使许多其他 NLP 任务受益。

2. 我们设计了一种新的事件检测方法。通过有效地去耦, 独立学习和自适应地融合事件表示, 我们的方法在稀疏和密集标记的触发器上都可以很好地工作, 并且在 ACE2005 和 KBP2017 数据集上都达到了最新的性能。

2 解耦用于事件检测的特定于词汇的表示形式和不依赖于词汇的表示形式

为了提炼歧视知识和泛化知识, 本节将事件表示法与

¹ 我们的源代码可在以下位置公开获得 <https://www.吉瑟布的.com / luyaojie / Δ-学习换编>

² https://Calcord.LDCUPEN.EDU/LDC2005_6T06

³ <https://TAC.Nist.GOV/2017/KBP/DATA.HTML>

分为三部分： $r_{\text{预计起飞时间}} r_w$ 电子束 $r_{\text{EB}} r_g$,

其中 r_w 是触发词的词表示形式，例如词嵌入 / ELMo（请注意， r_w 在我们所有的培训过程中都是固定的）；

r_d 是词法特定的事件表示形式，它

捕获歧视知识； r_g 是一种无词法表示形式，可以捕获泛化知识。通过解耦事件表示， r_d 和 r_g 将使用本节中的 Δ 学习算法 3。最后，提出了门机制来自适应地融合以上表示以进行事件检测。

形式上，事件检测实例是一对触发器候选及其上下文，即 $x = (t, c)$ ，其中 t 是触发器候选，而 $c =$

$\{c_m, \dots, c_{-1}, c_1, \dots, c_m\}$ 是其上下文。考试-

ple（被解雇，“那名官员是他的工作。”）是候选人被解雇的实例。

遵循先前的工作（阮和格里什 男人, 2015; 刘等., 2018a），给定实例 x ，我们将每个令牌 t_i 嵌入为 $t_i = [p_w; \text{聚丙烯}; p_e]$ ，其中 p_w 是其单词嵌入， p_p 是其位置嵌入， p_e 是其实体标签嵌入。因此， t_0 是触发候选的表示。在本文中，词汇专用模型 Θ_d 和无词汇模型 Θ_g 使用独立的嵌入。

2.1 词法表示

特定于词汇的表示法旨在捕获判别信息，以区分歧义触发词。例如，我们要

捕获 {官员, 工作, ...} 线索的表示形式用于将 (发射, EndPosition) 与 (发射, 攻击) 进行区分，以及 {坦克, 士兵, ...} 用于区分 (触发, 终止) 发出 (触发)。

为了捕获触发器候选者的判别线索，我们设计了一个以词汇为中心的上下文选择注意。我们将其称为 ATT-RNN 并描述如下。

以词法为中心的上下文选择。为了选择判别性上下文词，关注性上下文选择机制对触发候选及其上下文词之间的关联进行建模。例如，我们希望我们的注意力机制能够捕获“工作”与 S_1 中触发的关联，以及“坦克”与 S_2 中触发的关联。

word) 对与多层感知器 (MLP) 的相关性，并使用 softmax 函数将相关性得分标准化为注意力权重：

$$\alpha_i = \frac{\exp(\text{MLP}([h_0; h_i]))}{\exp(\text{MLP}([h_0; h_+]))} \quad (1)$$

具体来说，我们首先将 $[t_m, \dots, t_0, \dots, t_m]$ 馈入双向 GRU，以获取所有令牌的上下文感知令牌编码 $[h_m, \dots, h_0, \dots, h_m]$ 。然后我们的注意力机制模型（触发，上下文

$$j \in c \quad 0 \leq j$$

给定注意权重，特定于词法的上下文表示总结为： $c_0 =$

$\sum_{i \in C} \alpha_i \cdot \text{嗨}$ 。最后是词法特定的代表

实例 x 的反感是串联

它的记号表示 h_0 和特定于词法的上下文表示 c_0 ，即 $r_d = [h_0; c_0]$ 。

特定于词汇的表示形式可以通过捕获（触发，上下文词）关联来有效地消除触发词的歧义。但是，这种表示是特定于词汇的，因此很难很好地概括为稀疏/看不见的单词。

2.2 无词法表示

与词法特定表示相反，无词法事件表示 rg 旨在捕获 ED 的泛化知识，该知识可以在不同的触发词之间传递。例如，我们想要捕获与触发词无关的知识，例如 “[Trigger] to death”（致命死亡），这是 Attack 事件的强大触发模式，可用于检测许多不同的触发词，例如被触发，被黑客入侵，击败。以此方式，通过利用这种知识，即使是看不见的触发候选者 t 也可以容易地被识别。

显然，无词法事件表示 rg 应该与词法无关，但特定于事件。为此，我们将 x 中的所有标记表示为 t_i ，然后为 rg 采用独立于词法的上下文选择模块。我们只使用 DMCNN (Chen 等., 2015) 作为我们与词法无关的上下文选择模块，但在本节中设计了一种新的敌对 Δ 学习算法 3.2 可以从 rg 中消除词法相关信息。

与词法无关的上下文选择。为了选择词法无关但事件相关的上下文词，我们采用与 Chen 等. (2015). 例如，我们要捕获与 S5 中的 Attack 事件相关的“死亡”和“刑事”。

给定令牌序列 $[t_{-m}, \dots, t_0, \dots, t_m]$ ，宽度卷积层从 $t_{i:i+h-1}$ 捕获局部上下文特征 l_i ： $l_i = \tanh(w \bullet t_{i:i+h-1} + b)$

b)，其中 w 是卷积滤波器， b 是偏置项。为了总结来自句子不同部分的重要信号，动态池化层 (Chen 等., 2015) 用于产生左侧

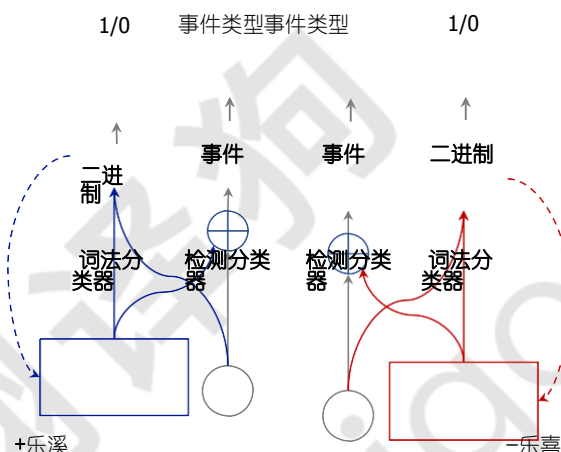
和右上下文特征 $1_{\text{剩}}^{\text{下}}$, $1_{\text{对}}^{\text{对}}$:

$$1_{\text{剩}}^{\text{下}} = \text{最大值 } 1_j, 1_{\text{对}}^{\text{对}} = \text{最大值 } 1_j \quad (2)$$

$$j < 0$$

$$j \geq 0$$

最后, 我们串联左上下文特征 $1_{\text{剩}}^{\text{下}}$ 和正确的上下文特征 $1_{\text{对}}^{\text{对}}$ 作为我们的无词汇表示法 $rg = [1_{\text{剩}}^{\text{下}}; 1_{\text{对}}^{\text{对}}]$.



2.3 表示融合的词汇门机制

以上两种表示是互补的

词法表征学习

无词汇表示学习

亚历克西斯

亚历克西斯

(t, c)

t/w

t/w

(t, c)

(a) 词汇增强型

(b) 词汇对抗性

彼此: r_d 捕获歧视知识-

边缘, 而 rg 则捕获通用知识。

但是, 简单的串联对于事件检测无效: 对于训练数据中频繁标记的触发词, 特定于词法的表示更有用; 对于标记稀疏或看不见的触发词, 无词法表示会有帮助。基于此观察, 我们的系统

t_{em} 需要更多地依赖 r_d 来检测频繁

图 3: 我们的 Δ 学习算法的框架。

每种事件类型 y_t 的概率 (包括非触发器的 NIL) :

$$p(y_t | x) = \frac{\exp(w_t = r_{\text{预计起飞时间}})}{\sum_{t=1}^T \exp(w_t = r_{\text{预计起飞时间}}) + b_t} \quad (6)$$

候选人被解雇, 但更多地举报 rg 以检测被入侵的 00V 候选人。也就是说, 我们需要适应性地

为不同的单词融合不同的表示形式, 而不是简单地将它们串联起来。

为了适应性地融合词汇专用表示 r_d , 无词汇表示 rg 和单词表示 rw , 我们设计了一种词汇门机制来融合不同的表示: $r_{\text{预计起飞时间}} = r_d \oplus rg$, 其中 \oplus 是融合门, 并且

$r_{\text{预计起飞时间}}$ 是最终事件表示。具体来说, 我

们首先将这些表示映射到通用空间:

$$r_d = f_{\text{规格} \rightarrow U}(r_d)$$

其中 \mathbf{w}_t 是权重向量, \mathbf{b}_t 是偏差术语。通过这种方式, 我们可以识别所有预定义事件类型的触发词。

3 通过提炼歧视和普遍化知识

$$\begin{aligned} R \ 7) \ g = FF \ REE \ \varepsilon \ u \\ (Rg) \end{aligned} \quad (3)$$

$$\begin{aligned} R \ 9) \ W = FRUCI \neg U \\ (RW) \end{aligned}$$

其中 $fSpec \rightarrow U(\bullet)$, $fFree \rightarrow U(\bullet)$ 和 $fLexi \rightarrow U(\bullet)$

是具有非线性函数的线性层; 然后通过门控机制将它们融合:

$$G_i = F(g(R_i)), \quad i \in \{d, g, w\}$$

$$g^i = \frac{\text{EXP}(g^i)}{\sum_{j \in \{d, g, w\}} \text{EXP}(g^j)} \quad (4)$$

g^i ($i \in \{d, g, w\}$) 分别表示 r^i s, r 如果和 r 金正日提供的证据的置信度; g^i 和 $g^{\sim i}$ 具有与 r^i 相同的尺寸; $fU \rightarrow G(\bullet)$ 是具有非线性函数的线性层

灰。最后, 我们结合所有表示形式:

$$\mathbf{r}_{\text{预计起飞时间}} = g_d \mathbf{0} \quad \mathbf{r}_{\text{身份证件}} = g_g \mathbf{0} \quad \mathbf{r}_g^i = g_w \mathbf{0} \quad (5)$$

其中 $\mathbf{0}$ 是逐元素乘法。

融合后, $\mathbf{r}_{\text{预计起飞时间}}$ 将被馈送到事件检测分类器, 该分类器将计算分类

Δ -学习

本节介绍了我们的 Δ 学习框架, 该框架可以独立学习词汇专用表示 r_d 和无词汇表示 rg 。要将歧视知识提炼为 r_d , 我们

签署词汇增强型 Δ 学习算法。至将广义知识提炼为 rg , 我们设计了一种词法对抗性 Δ 学习算法。最后,

我们微调图 1 中的完整事件检测模型 2.

3.1 通过提炼歧视知识

词汇增强型 Δ 学习

本节介绍了词汇增强型 Δ -词法表示 r_d 的学习算法。为了确保 r_d 与事件相关且与词法相关, 我们使用两种类型的监督信号: 首先, 我们希望学习的表示形式 r_d 可以借助 word 来预测其事件类型 y 。代表 rw ; 第二, 我们要

学习 r_d 也可以预测其触发词 t 。对于

例如, 我们希望实例的学习到的 r_d (被发射, 一个死者) 可以预测其事件类型 Attack 和触发词被发射。

为了实现上述目标，我们删除了图中的无词法部分 2 并在图中显示了词法增强的 Δ 学习框架 3 (a). 词汇增强型学习框架的输入是三元组 (t, c, w) ，其中 t 是触发器， c 是上下文， w 是样本单词。输出结果是双重的：事件分类器将输出事件类型 (t, c) ，如果 $t = w$ ，则辅助词法分类器将输出 1，否则输出 0。这样，事件分类器可以将事件类型监督信号传播到我们的特定于词法的表示学习组件，并且辅助词法二进制分类器确保所学习的表示法 r_d 是特定于词法的。

具体来说，对于每个 ED 实例 $x = (t, c)$ ，我们生成带有标签 $(y, 1)$ 的正词汇增强训练实例 (t, c, t) 和 n 个负实例⁴ (t, c, w) 带有标签 $(y, 0)$ ，其中 w 是从上下文 c 随机采样的单词。

对于火车数据集 D 中的每个 ED 实例 $x = (t, c)$ ，事件分类器损失为：

$$L_{event} = - \sum_{(x_k, y_k) \in D} \log p(y_k | x_k) \quad (7)$$

词汇二分类器的损失为：

$$L_{vocab} = - \sum_{(x_k, t) \in D} \log p(t | x_k) - \sum_{j=1}^n \log p(0 | x_k, w_{kj}) \quad (8)$$

因此，词汇增强的损失函数

Δ 学习是：

$$L_{增强} = L_{事件} + L_{词汇} \quad (9)$$

通过添加辅助词法分类任务，此学习算法将确保学习的表示形式既与事件相关，又与词法相关—相关。

3.2 通过词法-学习-学习来推广泛化知识

与词汇专用表示法 r_d 相比，无词汇表示法 r_g

也就是说，我们的对抗性 Δ 学习框架的输入和输出仍然是 (t, c, w) 和 $(y, 1/0)$ 。事件分类器用于将事件类型监督信号传播到我们的无词汇表示学习组件，以便 r_g 捕获事件相关信息。图之间的区别 3 (a) 和 3 (b) 他们使用不同的辅助任务：3 (a) 使用词汇增强的辅助任务，并且 3 (b) 使用词汇对抗辅助任务。

为了消除特定于词汇的信息，我们设计了一个两人最小/最大游戏 (好人等., 2014) 用于词汇对抗辅助任务。给定 (t, c, w) ，我们的二进制词法分类器 Θ_{DeLexi} 试图预测 r_g 是否特定于 w ，但是无词法模型 Θ_g 试图产生 r_g 来混淆 Θ_{DeLexi} 。词汇对抗 Δ 学习的最小-最大目标函数为：

$$L_{MinMAX} = \min_{\Theta_{DeLexi}} \max_{\Theta_g} \sum_{(x_k, t) \in D} \log p(t | x_k) - \sum_{j=1}^n \log p(0 | x_k, w_{kj}) \quad (10)$$

$$\hat{\theta} = \frac{\text{最小}}{\text{最大}} \frac{\text{最大}}{\text{最小}} \theta$$

这样，我们可以从 r_g 中删除词法特定的信息。

上述对抗性损失导致两个不同

Θ_g 和 Θ_{DeLexi} 的优化方向，其中

需要消除词汇专用信息，以便可以在不同单词之间进行传递。为了实现这一目标，我们采用对抗技术并设计了词汇对抗 Δ 学习算法。

具体来说，我们删除了图中的词法特定部分 2 并在图中显示了词汇对抗 Δ 学习框架 3 (b). 我们可以看到

⁴在本文中，我们将 n 设置为 1。

可以通过梯度反转层实现 (Ganin 等., 2016) 进行反向传播。也就是说, $\mathcal{L}_{\text{main}}$ 与主要的 ED 任务目标 $\mathcal{L}_{\text{event}}$ 共同进行了优化, 而 $\mathcal{L}_{\text{adver}}$ 当达到 reg 时, 唾液损失以 $\lambda_{\text{进阶}}$ 因子逆转。通过这种方式, 我们可以统一这些组件的优化方向。因此, 我们的词汇对抗性的损失函数

Δ 学习是:

$$\mathcal{L}_{\text{对手}} = \mathcal{L}_{\text{事件}} + \lambda_{\text{进阶}} \mathcal{L}_{\text{最小}} \quad (11)$$

以下刘等. (2019), 我们将词汇对抗 Δ 学习分为两个阶段:

1. 在预训练阶段, 我们首先使用主要 ED 任务目标更新 Θ_g , 然后冻结 Θ_g 并使用等式更新 Θ_{DeLexi} .
2. 在对抗学习阶段, 我们使用方程式更新参数 11.

在实践中, 我们发现因子 $\lambda_{\text{进阶}}$ 对最小-最大博弈的偶数敏感。较大的 $\lambda_{\text{进阶}}$ 容易使二进制词法分类器变弱 (二进制分类精度趋于

50%). 在本文中, 将 $\lambda_{进阶}$ 设置为 1^{-3} , 并且将在对抗学习阶段, 我们的二进制词汇分类器 Θ_{DeLexi} 的准确性始终保持 75% 以上。

通过添加辅助词法对抗任务, 该学习算法将确保学习的表示形式与事件相关但与词法无关。

3.3 全模型微调

给定预训练的词汇专用表示模型 Θ_d 和预训练的无词汇表示模型 Θ_g , 我们最终微调了图 9 中的完整模型 2 通过优化事件分类损失函数:

$$L(\cdot) = Leunt + \lambda_{reg} \cdot R_g \cdot 1 \times 1 \quad (12)$$

其中, λ_{reg} 是正则项的权重系数, 而 Θ 表示所有参数。L (Θ) 可以使用基于小批量的随机梯度下降算法 (例如 Adadelta (蔡勒, 2012)).

4 个实验

4.1 实验设置

数据集。我们在两个标准的英语事件检测数据集上进行实验: ACE2005 和 KBP2017。

ACE2005 (LDC2006T06) 包含 599 个文档, 其中注释了 33 个事件类型。之前的研究 (廖和格里什曼, 2010; 里等., 2013; Chen 等., 2015; 刘等., 2017, 2018a), 我们在实验中使用相同的 529/30/40 train / dev / test 文档拆分。与以前的研究一样, 我们使用 ACE2005 作为主要数据集 (阮和格里什曼, 2018)。

KBP2017 (LDC2017E55) 包含 500 个带有 RichERE 批注的文档, 用于 TAC KBP 2017 评估。对于模型训练, 我们使用以前注释的 RichERE 数据集, 包括 LDC2015E29, LDC2015E68, LDC2016E31 和 TAC KBP 2015-2016 评估数据集。遵循先前的工作 (Lin 等., 2018a), 我们从 2016 年评估数据集中随机抽取了 20 个文档作为开发集。

我们使用精度, 召回率和 F1 分数评估不同的事件检测系统。对于 ACE2005, 我们计算这些条件的方式与以前的工作相同 (Li 等., 2013; Chen 等., 2015)。对于 KBP2017, 因为 TAC KBP2017 al-

要求每个团队提交 3 次不同的运行, 以使我们的结果与评估结果具有可比性,

我们在开发集中选择每个系统的 3 个最佳运行, 并使用官方评估工具包报告其中的最佳测试性能⁵, 在先前的工作中被称为 Best3 (林等., 2018a)。

基线。我们将我们的方法与三种基准进行比较:

基于功能的方法依赖于丰富的手工设计功能, 包括: MaxEnt (Li 等., 2013), 采用手工设计的功能, 并使用最大熵分类器; 合并的 PSL (刘等., 2016b) - 报告得最好的基于功能的系统, 它使用概率软件逻辑框架结合了全局和潜在功能。

基于表示学习的方法采用神经网络自动提取事件检测功能, 包括: DMCNN (Chen 等., 2015) 使用 CNN 作为句子特征提取器, 并将句子特征和词法特征连接起来进行事件检测分类器; NC-CNN (NC-CNN) 阮和格里什曼, 2016) 通过对跳跃图进行建模以扩展非连续 k-gram 来扩展传统的 CNN; 双 RNN (Nguyen 等., 2016), 它使用双向 RNN 特征提取器的附加依赖项特征嵌入每个令牌, 并联合提取带有其参数的触发器。

基于外部资源的方法旨在利用外部资源来增强事件检测, 包括: SA-ANN-Arg (刘等., 2017) 通过监督注意力机制注入事件参数信息; GCN-ED 阮和格里什曼, 2018) 利用图卷积网络 (GCN) 利用句法信息来捕获更准确的上下文; 格拉姆特刘等., 2018a) 利用多语言信息进行更准确的上下文建模; 乙型肝炎病毒表面抗原 Chen 等., 2018), 它融合了句子级和文档级信息, 并共同检测句子中的不同事件。

对于我们的方法和所有基准, 我们采用使用 Skip-gram 进行预训练的词嵌入⁶和公开发布的 ELMo 模型⁷。我们还报告了 ELMo 的性能, 作为演示通用预训练表示形式性能的基准。所有超参数都根据开发集进行了调整。

⁵<https://github.com/hunterhector/EvmEval>

⁶:[HTTPS://CODE. GoGoLe. COM/SaveV/P/Word2VEC](https://code.google.com/SaveV/P/Word2VEC/)⁷:[HTTPS://AelnLP. Org/ELMO](https://AelnLP.Org/ELMO)

	P	R	F1
基于特征的方法			
最大熵	74.5	59.1	65.9
联合 PSL	75.3	64.4	69.4
基于表示学习的方法			
德姆 cnn	75.6	63.6	69.1
碧 RNN	66.0	73.0	69.3
数控美国有线电视新闻网	-	-	71.3
基于外部资源的方法			
SA-ANN-Arg (+参数)	78.0	66.3	71.7
GMLATT (+多国语言)	78.9	66.9	72.4
GCN-ED (+语法)	77.9	68.8	73.1
HBTNGMA (+文档)	77.9	69.1	73.3
我们的方法			
埃尔莫	75.6	62.3	68.3
Δ_{w2v}	71.8	70.8	71.3
$\Delta_{\text{的}}$	73.7	71.9	72.8
Δ_{w2v}	74.0	70.5	72.2
榆树	76.3	71.9	74.0

表 2: 在 ACE 2005 上的实验结果。为了公平比较, 基线结果改编自其原始论文。

4.2 整体表现

表 2 显示了所有基准和我们的方法的 ACE2005 总体结果。对于我们的方法, 我们显示了四种设置的结果: 我们使用词嵌入作为其词表示形式 $rw - \Delta_{w2v}$ 的方法; 我们使用 ELMo 作为 rw 的方法- Δ_{ELMo} ; 我们的方法只是将 $[r_d, rg, rw]$ 连接为实例表示- $\Delta_{\text{康塔特}}$ 。从表 2, 我们可以看到:

1. 通过提炼歧视知识和概括知识, 我们的方法获得了最先进的性能。与最佳功能系统相比, Δ_{w2v} 和 Δ_{ELMo} 增益

F1 评分 2.8 和 4.6 的改进。相比

基于表示学习的基线

Δ_{w2v} 和 Δ_{ELMo} 的性能均优于所有。值得注意的是, 使用外部资源, Δ_{ELMo} 优于所有基准。

2. 通过逐步提取归纳知识, 我们的方法可以实现较高的查全率和高精度。我们的方法

	P	R	F1
--	---	---	----

TAC 2017 ED 曲目前三名

TAC 2017 第三名 54.27 46.59 50.14

获得了很高的召回率- 71.9, 大大优于大多数方法, 并保持了较高的精度

- 76.3. 我们认为这是因为泛化知识是使用 Δ -学习逐步提炼出来的, 因此在训练过程中无需进行精确召回权衡。

3. 词汇门为自适应融合判别提供了一种有效的机制,

TAC 2017 第二名	52.16	48.71	50.37
TAC 2017 第一名	56.83	55.57	56.19

我们的方法

Δ_{w2v}	62.84	50.36	55.91
----------------	--------------	-------	-------

<u>榆树</u>	62.30	53.77	57.72
-----------	-------	-------	--------------

表 3: 在 TAC KBP 2017 评估数据集上的实验结果。

归纳和概括知识。与天真的融合基线相比， $\Delta_{\text{康塔特}}$ ， Δ_{w2v} 和 Δ_{ELMo} 分别提高了 0.9 和 1.2 F1。这意味着适应性融合机制既可以从区分知识和泛化知识中受益，又可以在两者之间进行权衡。

4. 尽管通用的预训练表示形式可以实现良好的性能，但特定任务的表示形式仍然至关重要。与强大的通用表示基线 ELMo 相比，我们针对特定任务的事件检测表示均实现了显著的性能提升。这也验证了

Δ 学习是增量学习特定于任务的表示的有效方法。

表 3 进一步将我们的方法与 TAC 2017 Event Detection Track 中的前 3 个系统进行了比较 (Mitamura 等., 2017). 由于这些团队在评估期间无法访问黄金实体信息，因此为了公平比较，我们在 KBP2017 实验中排除了实体嵌入。我们可以看到，尽管这些方法是利用各种外部资源的集成模型，但所提出的方法可以大大胜过 TAC 2017 中最好的 ED 系统。

4.3 详细分析

要详细分析我们的方法的效果，请参见表 4 展示了我们的方法在不同类型的触发字上的性能，包括：

OOV (无语) 和 OOL (无语) 与表中的相同

1.

稀疏实例意味着给定单词 $P(e | w) = \frac{\#(e, w)}{\#(w)}$ 的事件触发率小于 10%

在训练语料库中，即出现词 w 的发生率 $< 10\%$ 标记为事件类型 e (包括 NIL)。

密集表示除 OOV 外的所有其他实例，

$\#(w)$

00L 和稀疏。

让电子束为我们的词汇门，表 4 显示以下设置的结果： $r_{EB}rw$, rg 和 $r_{EB}rg$ (即我们的完整模型)。为了妖魔表 Δ 学习的效果，表 4 还显示了非 Δ 学习版本的结果 $r_{EB}rw$, rg 和 $r_{EB}rg$ 。在这组中

对 $ting$, r_d 和 rg 进行训练，而无需使用辅助任务。从表 4，我们可以看到：

1. 先前的方法在提取歧视知识方面效果很好，但在提取泛化知识方面效果不佳。先前的方法在 Dense 实例上获得较高的 F1 分数，但是在稀疏标签实例上的性能却很差。任务特定表示 (ATT-RNN 和 DMCNN) 仅与通用单词表示 rw (ELMo) 取得相似的性能。

2. Δ 学习对于逐步提取知识有效。与非 Δ 学习版本 rg 和 $r_{EB}rg$ 可以提炼一般知识化，即获得 8.5、12.3 和 9.8 F1 对 00V、00L 和稀疏实例的改进。 $r_{EB}rw$ 可以提取出更多的判别式知识，而不是非 Δ 学习版本。

3. 分解策略，即学习和融合独立知识非常有效。

对表征学习有效。通过分解， rg 和 $r_{EB}rg$ 可以捕获泛化知识， $r_{EB}rw$ 可以捕获判别式知识，它们是相辅相成的。从 rw 开始，我们的方法可以通过以下方式逐步提取 r_d 和 rg 中的知识 Δ 学习。通过融合独立知识通过有效的词汇门在 r_d 和 rg 中的边， $r_{EB}rg$ 和 rw 在 00V、00L 和 Dense 实例。

5 相关工作

事件检测。近年来，神经方法在事件检测中取得了重大进展。大多数神经方法都专注于学习有效的实例表示 (陈等., 2015; 阮和格里什曼, 2015, 2016; Nguyen 等., 2016; 冯等., 2016; 盖尼等., 2016; Lin 等., 2018b)。这些方法的主要缺点是，它们大多只学习单个词法和特定于词法的表示形式，这在提取歧视知识时效果很好，而在归纳知识方面却效果很差。

一些方法可以增强代表性学习-

交涉	00V	00L	稀疏	稠密
ATT-RNN	38.7	6.2	36.7	77.7
DMCNN	32.4	9.0	43.1	77.6
埃尔莫	31.3	9.0	47.1	78.0
r_d rg rw (w /	40.0	8.8	50.0	78.7
rg rw (不带	47.1	11.1	54.6	78.8
rd rg rw	40.0	11.4	52.8	78.8
rg rw	32.3	12.3	43.1	79.1
rd rg rw	55.6	23.4	64.4	78.2
rd rg rw	57.4	26.7	55.6	80.0

表 4：在不同类型的触发词上，不同表示形式 (ELMo 作为单词表示形式 rw) 的结果 (F1 分数)。为了公平起见，与标准 DMCNN (Chen 等., 2015) 在表格中 1

和表 2，DMCNN*排除词法功能，但在-包含实体功能。

使用外部资源。一种策略是运用额外的知识来更好地学习表示，例如文档 (段等., 2017; Chen 等., 2018; 刘等., 2018b)，句法信息 (阮和格里什曼, 2018; 沙等., 2018; 奥尔等., 2018; 刘等., 2018c)，事件参数 (刘等., 2017)，知识库 (杨和米切尔, 2017; 卢和阮, 2018) 和多语言信息 (刘等., 2018a)。另一种策略是从额外的知识库中生成其他培训实例 (刘等., 2016a; Chen 等., 2017) 或新闻段落群集 (弗格森等., 2018)。我们的方法不使用任何外部资源，可以很好地补充这些方法。

通过辅助学习进行表征学习。近年来，已经提出了许多辅助学习技术，以实现更好的表示学习。自我监督学习通过设计辅助任务而不是使用手动标记的数据来学习表示形式。示例包括视觉任务中的着色 (多尔施和 齐塞曼, 2017)，文本任务中的语言建模 (丽, 2017)。对抗性学习试图通过恶意输入来欺骗模型 (Kurakin 等., 2016)，它已广泛用于许多场景中，例如领域适应 (Zeng 等., 2018)，知识蒸馏 (秦等., 2017) 和属性清理 (埃拉扎尔和戈德堡, 2018)。

一些基于对抗的技术已用于事件检测。Hong 等. (2018) 通过自我调节克服了训练期间的虚假特征。刘等. (2019) 使用师生网络从外部 NLP 资源中提取额外的知识。本文采用 ad-

全方位 Δ 学习算法, 可消除事件表示中的词汇信息, 从而可以逐步提炼判别和概括性知识。

6 结论

本文提出了一种新的表示学习框架- Δ 学习, 它可以提炼歧视知识和泛化知识用于事件检测。具体来说, 两个有效提出了 Δ 学习算法来独立地提取歧视和概括性知识, 并设计了词汇门机制来自适应地融合不同的知识。实验结果证明了我们方法的有效性。表示学习是 NLP 任务的一项基本技术, 特别是解决自然语言表达的歧义和多样性问题。为了将来的工作, 我们计划使用我们的 Δ 学习框架研究新的辅助 Δ 学习算法。

致谢

我们衷心感谢审稿人的宝贵意见。此外, 这项工作得到了中国国家重点研究计划的资助 (2018YFB1005100); 国家自然科学基金资助号: 61433015、61572477 和 61772505; 和青年精英科学家资助计划 YESS20160177。

参考

- 大卫·安 2006 年。事件提取的阶段. 在关于时间和事件的注释和推理讲习班的会议记录中, 第 1-8 页, 澳大利亚悉尼。计算语言学协会。
- 詹姆斯·艾伦。2012。主题检测和跟踪: 基于事件的信息组织, 第 12 卷。Springer 科学与商业媒体。
- 纳塔奈尔·钱伯斯 (Nathanael Chambers) 和丹·尤拉夫斯基 (Dan Jurafsky)。2008 年。云水-指导性学习叙事事件链. 在 ACL-08: HLT 会议记录中, 第 789-797 页。计算语言学协会。
- 陈玉波, 刘树林, 张翔, 刘康和赵军。2017 年。自动标记的数据生成-大规模事件提取. 在计算语言学协会第 55 届年度会议论文集 (第 1 卷: 长论文) 中, 第 409-419 页, 加拿大温哥华。计算语言学协会。

陈玉波, 徐立恒, 刘康, 曾道建和赵军。2015 年。通过 dy-提取事件 纳粹多池卷积神经网络. 在计算语言学协会第 53 届年会和第七届国际自然语言处理联合会议 (第 1 卷: 长论文) 的会议记录中, 第 167-176 页, 中国北京。计算语言学协会。

陈玉波, 杨航, 刘康, 赵军和贾彦涛。2018。通过 hier-进行集体事件检测 带有门控多级别注意机制. 在《2018 年自然语言处理中的经验方法会议》上, 第 1267-1276 页。计算语言学协会。

卡尔·多尔施 (Carl Doersch) 和安德鲁·齐瑟曼 (Andrew Zisserman)。2017 年。多-任务自我监督的视觉学习. 在 IEEE 国际计算机视觉会议 (ICCV) 中, 第 2051-2060 页。

段少阳, 何瑞芳和赵雯丽。2017 年。利用文档级别信息进行改进 通过递归神经网络进行事件检测. 在第八届国际自然语言处理联合会议论文集 (第 1 卷: 长论文) 中, 第 352-361 页。亚洲自然语言处理联合会。

Yanai Elazar 和 Yoav Goldberg。2018。对抗性从文本数据中删除人口统计属性. 在《2018 年自然语言处理中的经验方法会议》上, 第 11-21 页。计算语言学协会。

冯小成, 黄立夫, 唐笃玉, 衡基, 秦兵和刘婷。2016 年一种语言-独立的神经网络用于事件检测. 在《计算语言学协会第 54 届年会论文集》(第 2 卷: 简短论文), 第 66-71 页中。计算语言学协会。

詹姆斯·弗格森 (James Ferguson), 科林·洛克德 (Colin Lockard), 丹尼尔·韦德 (Daniel Weld) 和汉纳内·哈吉希尔 (Hannaneh Hajishirzi)。2018。半监督事件用复述词簇提取. 在计算语言学协会北美分会 2018 年会议论文集: 人类语言技术, 第 2 卷 (短论文), NAACL'2018, 第 359-364 页。计算语言学协会。

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, Francois Laviolette, Mario Marchand 和 Victor Lempitsky。2016 年神经网络的领域对抗训练. J. 马赫学习. Res., 17 (1): 2096-2030。

雷扎·加埃尼 (Reza Ghaeini), 小莉·弗恩 (Xiaoli Fern), 黄亮和普拉萨德·塔德帕利 (Prasad Tadepalli) 2016 年事件块检测 前后递归神经网络. 计算语言学协会第 54 届年会论文集 (第 2 卷:

简短论文), 第 369-373 页, 德国柏林。计算语言学协会。

的第二十三届 AAAI 人工智能会议上, 第 4865-4872 页, 美国纽约。人工智能促进协会。

伊恩·古德费洛 (Ian Goodfellow), 让·普格特·阿巴迪 (Jean Pouget-Abadie), 梅赫迪·米尔扎 (Mehdi Mirza), 许冰, 大卫·沃德·法利 (David Warde-Farley), 谢吉·奥扎尔 (Sherjil Ozair), 亚伦·库维尔 (Aaron Courville) 和 Yoshua Bengio。2014 年生成广告通用网。《神经信息处理系统的发展》, 第 2672 至 2680 页。Curran Associates, Inc.

于洪, 周文轩, 张静丽, 朱巧明和周国栋。2018。自我调节: 利用生成对抗网络来证明事件检测。计算语言学协会第 56 届年会论文集, 第 515-526 页。计算语言学协会。

Alexey Kurakin, Ian J. Goodfellow 和 Samy Bengio。2016 年大规模对抗机器学习。CoRR, abs / 1611.01236。

齐力, 亨吉和梁煌。2013。联合活动通过具有全局特征的结构化预测进行提取功能。在计算语言学协会第 51 届年会论文集 (第 1 卷: 长论文) 中, 第 73-82 页, 保加利亚索菲亚。计算语言学协会。

廖莎莎和拉尔夫·格里什曼。2010。使用文档注释级别的跨事件推理以改善事件萃取。在《计算语言学协会第四十八届年会论文集》, 第 789-797 页。计算语言学协会。

林宏宇, 陆耀杰, 韩先培和孙乐。2018A。自适应缩放, 用于稀疏检测地层提取。计算语言学协会第 56 届年会论文集, 第 1033-1043 页。计算语言学协会。

林宏宇, 陆耀杰, 韩先培和孙乐。2018B。中国活动的掘金提案网络发现。计算语言学协会第 56 届年会论文集, 第 1565-1574 页。计算语言学协会。

刘健, 陈玉波和刘康。2019。利用-真相: 基于对抗的模仿知识蒸馏方法用于事件检测。在 AAAI 第三十三届 AAAI 人工智能会议上, 2019 年。人工智能促进协会。

刘健, 陈玉波, 刘康和赵军。2018A。通过门控多语言注意力进行事件检测机制。在 AAAI' 2018 年

刘少波, 成瑞, 余晓明和成学琪。2018B。利用上下文信息 通过动态内存网络进行事件检测. 在《 2018 年自然语言处理中的经验方法会议》上, 第 1030-1035 页。计算语言学协会。

刘树林, 陈玉波, 何世柱, 刘康和赵军。2016a。利用框架网络进行改进 自动事件检测. 在《计算语言学协会第 54 届年会论文集》(第 1 卷: 长论文) 中, 第 2134-2143 页, 德国柏林。计算语言学协会。

刘树林, 陈玉波, 刘康和赵军。2017 年。利用论点信息来改善事件 通过监督注意力机制进行检测. 在计算语言学协会第 55 届年度会议论文集 (第 1 卷: 长论文) 中, 第 1789-1798 页, 加拿大温哥华。计算语言学协会。

刘树林, 刘康, 何世柱和赵军。2016B。基于概率软逻辑的方法来利用- 在事件分类中收集潜在的全局信息 阳离子. 在“第三十届 AAAI 人工智能会议论文集”中, AAAI'16, 第 2993-2999 页。AAAI 新闻。

刘小, 罗准臣和黄鹤雁。2018C。通过注意力联合提取多个事件 基于图的信息聚合. 在《 2018 年自然语言处理中的经验方法会议》上, 第 1247-1256 页。计算语言学协会。

卢维依和阮天恩。2018。相似但 不一样-词义歧义改善 通过神经表示匹配进行事件检测. 在《 2018 年自然语言处理中的经验方法会议》上, 第 4822-4828 页。计算语言学协会。

三村辉子, 刘正中和爱德华·H·霍维 (Eduard H Hovy)。2017 年。事件检测, 共指和安全问: 下一步是什么? tac kbp 概述 2017 年活动跟踪. 在 TAC。

Thien Huu Nguyen, Kyunghyun Cho 和 Ralph Grishman。2016 年通过循环提取联合事件 神经网络. 在计算语言学协会北美分会 2016 年会议论文集: 人类语言技术, 第 300-309 页, 加利福尼亚州圣地亚哥。计算语言学协会。

Thien Huu Nguyen 和 Ralph Grishman。2015 年。事件 卷积检测和域自适应 神经网络. 计算语言学协会第 53 届年会暨第 7 届国际联席会议论文集

自然语言处理 (第 2 卷: 简短论文), 第 365-371 页, 北京, 中国。计算语言学协会。

马修 ● D ● 泽勒。2012。Adadelta: 自适应学习-利率法. CoRR, abs / 1212.5701。

Thien Huu Nguyen 和 Ralph Grishman。2016 年国防部-用卷积事件检测的 Eling 跳过语法 国家神经网络. 在《2016 年自然语言处理中的经验方法会议》的会议记录中, 第 886-891 页, 德克萨斯州奥斯汀。计算语言学协会。

曾佳丽, 苏劲松, 温华庭, 刘洋, 谢军, 尹永京和赵建强。2018。带词的多域神经机器翻译 级别域上下文歧视. 在《2018 年自然语言处理中的经验方法会议》上, 第 447-457 页, 比利时布鲁塞尔。计算语言学协会。

Thien Huu Nguyen 和 Ralph Grishman。2018。图形 具有参数感知池的卷积网络 事件检测. 在 AAAI' 2018 年第三十二届 AAAI 人工智能会议上, 第 5900-5907 页, 纽约, 纽约, 美国。人工智能促进协会。

Walker Orr, Prasad Tadepalli 和 Xiaoli Fern。2018。使用神经网络进行事件检测: 严格 实证评估. 在《2018 年自然语言处理中的经验方法会议》上, 第 999-1004 页。计算语言学协会。

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee 和 Luke Zettlemoyer。2018。深度上下文词表达 判刑. 在《计算语言学协会北美分会 2018 年会议论文集: 人类语言技术》, 第 1 卷 (长论文), 第 2227 至 2237 页。计算语言学协会。

秦连辉, 张志松, 赵海, 胡志婷和邢志伟。2017 年。对抗结缔 利用网络隐性话语关系 分类. 在《计算语言学协会第 55 届年会论文集》(第 1 卷: 长论文) 中, 第 1006 至 1017 页, 加拿大温哥华。计算语言学协会。

Marek Rei。2017 年。半监督多任务学习- 序列标记. 在计算语言学协会第 55 届年度会议论文集 (第 1 卷: 长论文) 中, 第 2121 至 2130 页, 加拿大温哥华。计算语言学协会。

雷莎, 冯谦, 张宝宝和隋志芳。2018。联合提取 事件触发器和参数 依赖桥 rnn 和基于张量的 论证互动. 在 AAAI' 2018 年第三十二届 AAAI 人工智能会议上, 第 5916-5923 页, 美国纽约。人工智能促进协会。

杨碧珊和汤姆·米切尔。2017 年。利用 lstms 中的知识库, 用于改进机器读. 在《计算语言学协会第 55 届年会论文集》(第 1 卷: 长论文) 中, 第 1436-1446 页。计算语言学协会。