

基于神经潜变量模型的开放域事件提取

小刘^{1,2}和黄艳艳^{1,2}和张悦^{3,4*}

¹北京理工大学计算机科学与技术学院

²浙江实验室, 中国

{消瘤, hhy63} @ bit.edu.cn

³西湖大学工程学院

⁴西湖高等研究院先进技术研究所

You. Auth. Auth. W. A. C. C. C. C. C. C.

摘要

我们考虑开放域事件提取, 即从新闻集群中提取无约束类型的事件的任务。构建了一种新的潜变量神经模型, 该模型可扩展到非常大的语料库。收集并手动注释数据集, 并设计特定于任务的评估指标。结果表明, 与用于事件模式归纳的最新方法相比, 所提出的无监督模型提供了更好的性能。

1 介绍

从新闻文本中提取事件已引起很多研究关注。该任务通常由两个子任务组成, 即模式归纳, 用于提取为给定事件类型指定参数槽的事件模板 (钱伯斯, 2013; Cheung 等。 , 2013; Nguyen 等。 , 2015; Sha 等人。 , 2016; 黄等人。 , 2016; 安, 2017; 袁等人。 , 2018) 和事件提取, 用于从一条新闻中识别带有填充槽的事件 (Nguyen 等。 , 2016b; Sha 等人。 , 2018; 刘等人。 , 2018a; 陈等人。 , 2018, 2015; 丰等。 , 2016; Nguyen 和 Grishman, 2016; 刘等。 , 2018b). 以前的工作重点是根据一组预先指定的事件类型 (例如纵火, 攻击或地震) 从单个新闻文档中提取事件。

虽然对于从新闻中跟踪高度特定类型的事件非常有用, 但上述设置对于安全和金融市场中的决策制定可能相对较少, 这可能需要有关广泛覆盖, 细粒度和动态演变的事件类别的全面知识。此外, 鉴于不同的新闻机构可以报告相同的事件, 可以利用冗余来更好地提取事件。在本文中, 我们调查开放域

——通讯作者。

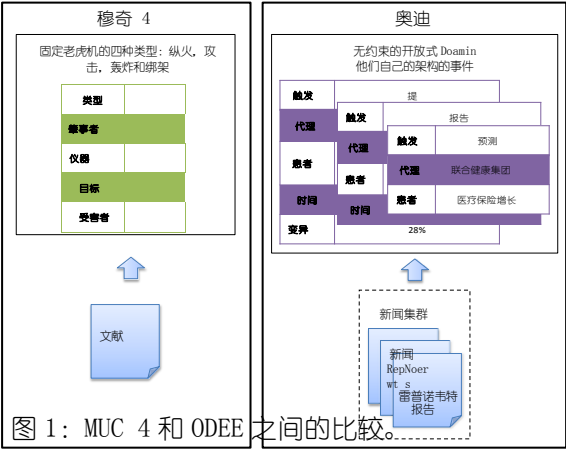


图 1: MUC 4 和 ODEE 之间的比较。

事件提取 (ODEE), 用于提取无约束类型的事件并从新闻报道集群中引入通用事件模式。

如图所示 1, 与 MUC 4 举例说明的传统事件提取任务相比 (桑德海姆, 1992), ODEE 的任务对建模提出了额外的挑战, 传统方法尚未考虑这些挑战。首先, 可以从新闻集群中提取多个事件, 其中事件可以灵活地在开放域中具有不同数量的插槽, 并且插槽可以是灵活的而没有相同的分布, 而不管事件类型如何。关于模式归纳的工作。其次, 应该考虑提及新闻集群中不同报告中的相同实体, 以提高绩效。

我们构建了一个无监督的生成模型来应对这些挑战。以前关于生成模式归纳的工作 (钱伯斯, 2013; Cheung 等。 , 2013; Nguyen 等。 , 2015) 依赖于手工制作的指标特征, 我们引入神经网络产生的潜在变量以获得更好的表现力。一种新颖的图模型

设计, 具有来自全局参数化正态分布的每个新闻集群的潜在事件类型向量, 以及实体的文本冗余特征。我们的模型利用了语境化的预训练语言模型 (ELMo, 彼得斯 等。(2018)) 和可扩展的神经变分推理 (斯里瓦斯塔瓦和萨顿, 2017)。

为了评估模型性能, 我们从 Google Business News 收集并注释了一个大型数据集¹ 具有不同的事件类型和可解释的事件模式。除了模式匹配的标准度量之外, 我们还基于 NPMI 调整时隙一致性 (Lau 等人., 2014) 用于定量测量时隙和模式的内在质量, 这些内在质量固有的集群。

结果表明, 我们的神经潜变量模型优于最先进的事件模式归纳方法。此外, 冗余对于改进开放域事件提取非常有用。学习参数的可视化表明我们的模型可以给出合理的潜在事件类型。据我们所知, 我们是第一个使用神经潜变量模型来诱导事件模式和提取事件的人。我们发布我们的代码和数据集 <https://github.com/LX867012528/ACL2019-ODEE>。

2 相关工作

最流行的模式归纳和事件提取任务设置是 MUC 4, 其中四种事件类型 - 纵火, 攻击, 轰炸和绑架

- 和四个插槽 - 犯罪者, 工具, 目标和受害者 - 已定义。我们比较了图中 MUC 4 和 ODEE 的任务设置¹。对于 MUC 4, 输入是单个新闻文档, 输出属于四种类型的事件, 其中模式由固定时隙组成。相反, 对于 ODEE, 输入是新闻集群而不是单个新闻, 输出是无约束类型的开放域事件和具有各种时隙组合的唯一模式。

Event Schema Induction 开创性工作研究模式 (Shinyama 和 Sekine, 2006; 斐乐 tova 等人., 2006; 邱等人, 2008) 和事件链 (钱伯斯和朱拉夫斯基, 2011) 用于模板诱导。对于 MUC 4, 目前的主导方法包括概率生成方法 (钱伯斯, 2013; Cheung 等., 2013; 阮 等., 2015) 联合模型谓词和 ar-

用于诱导时隙的 gument 赋值和 ad-hoc 聚类算法 (Sha 等人., 2016; 黄 等., 2016; 安, 2017; 袁等人., 2018)。这些方法都依赖于手工制作的离散特征, 而无需完全模拟文本冗余。还有使用神经语言模型建模事件模式和脚本的工作 (莫迪和 季托夫, 2014; 鲁丁格等人., 2015; Pichotta 和 穆尼, 2016), 但他们没有探索神经潜在变量和冗余。

事件提取工作通常假定给出了事件模式, 识别事件触发器及其相应的参数。这可以视为 ODEE 的子任务。现有工作利用句子级别 (McClosky 等., 2011; 里 等., 2013; 刘等人., 2016; 杨和米切尔, 2016) 和文件级统计 (廖和格 伊斯曼, 2010b; 吉和格里什曼, 2008; 香港 等., 2011; Reichart 和 Barzilay, 2012)。还有使用 RNN 的工作 (Nguyen 等., 2016b; Sha 等人., 2018; 刘等人., 2018a; 陈 等., 2018CNs) 陈等人., 2015; 冯等人., 2016; Nguyen 和 Grishman, 2016) 和 GCN (刘等人., 2018b) 表示事件的句子。事件提取已被视为监督或半监督 (廖和格里什曼, 2010a; 黄和 Riloff, 2012) 任务。相比之下, ODEE 是完全无监督的设置。

Tweet Streams 中的事件发现提取了新闻有价值的单词, 片段和帧集群。已经使用了监督和非监督方法。前者 (Sakaki 等., 2010; Benson 等人., 2011) 通常设计用于监视某些事件类型, 而后一个集群根据其突发性特征 (贝克尔等人., 2011; 崔等人., 2012; 李等人., 2012; 里特等人., 2012; 秦等人., 2013; Ifrim 等., 2014; 麦克明和何塞, 2015; 秦等人., 2017)。这一系列工作类似于我们使用信息冗余的工作, 但不同, 因为我们专注于正式的新闻文本和诱导结构事件模式。

第一故事检测 (FSD) 系统旨在识别讨论之前未报告的事件的新闻文章。大多数关于 FSD 的工作通过找到新文件的最近邻居来检测第一个故事 (库马兰和艾伦, 2005; 莫兰 等., 2016; Panagiotou 等人., 2016; Vuurens 和 德弗里斯, 2016)。这一系列工作利用大量流中的文本冗余来预测文档是否包含作为 clas 的新事件 -

¹<https://news.google.com/?hl=en-US&gl=美国与CEID=US>: 恩, 从 2018 年 10 月到 2019 年 1 月爬行。

资格任务。相反，我们研究事件模式并提取详细事件。

3 任务和数据

任务定义。在 ODEE 中，输入由新闻集群组成，每个集群包含有关同一事件的报告。输出一包开放域事件，每个事件由事件触发器和自己的模式中的事件参数列表组成。在大多数情况下，一个事件在语义上足以表示输出。

在形式上，给定一个包含一组新闻集群的开放域新闻语料库 $\{c \in N\}$ ，假设在新闻集群 c 中有 M_c 新闻报道 $\{d_i \in c \mid i = 1, \dots, M_c\}$ ，其中 d_i 是前面提到的一组开放域事件，TE 是一组模式。

为这组事件定义语义槽。数据采集。我们抓取 Google 商业新闻的新闻报道，该新闻报道提供有关来自不同来源的相同活动的新闻群组。在每个新闻集群中，不超过五个新闻报道。对于每个新闻报道，我们获取标题，发布时间戳，下载时间戳，源 URL 和全文。从 2018 年 10 月 17 日到 2019 年 1 月 22 日，我们总共获得了 5518 个商业新闻报道，其中包括 13,047 个新闻报道，共 288 批次。每天执行爬虫约三次。全文语料库作为 GNBUSINESS-Full-Text 发布。在本文中，我们通过保留标题和第一段，作为 GNBUSINESS-All 发布来修剪每个新闻集群中的新闻报道。

灵感来自 FrameNet 中的常规插槽 (贝克等人, 1998)，我们为开放域事件类型设计参考事件模式，其中包括八个可能的插槽：代理，患者，时间，地点，目标，旧值，新值和变体。Agent 和 Patient 分别是触发器的语义代理和患者；目标是事件的目标或原因。如果事件涉及值更改，则旧值为旧值提供，新值为新值提供，而“变量”为“新值”和“旧值”之间的变化。请注意，我们定义的角色更具主题性，对特定事件的特定性不如某些现有事件提取数据集那样 (晒-

海姆, 1992; Nguyen 等., 2016a)，因为我们想让我们的数据集一般而且对于 a 有用广泛的开放领域条件。我们为未来的工作留下更精细的角色类型。

分裂	#C	#R	#S	#W
测试	574	2,433	5,830	96,745
开发	106	414	991	16,839
未标记	12,305	52,464	127,416	2,101,558
所有	12,985	55,311	134,237	2,215,142
全文	12,985	55,311	1,450,336	31,103,698

表 1: 数据拆分统计。(C 新闻集群; R 新闻报道; S 句; W 字。)

数据集	#D	#L	#T	#S
穆奇 4	1700	400	4	4
王牌 2005	599	599	33	36
埃尔	562	562	38	27
阿斯特里	1038	100	12	18
全球商务	12,985	680	-	8

表 2: 与现有数据集的比较。(D 个文件或新闻集群; L 个标签文件或新闻集群; T 个事件类型; S 个插槽。)

我们随机选择了 18 批新闻集群，总共有 680 个集群，按照比例为 1: 5 将它们划分为开发集和测试集。开发集，测试集和其他未标记集群作为 GNBUSINESS-Dev 发布，GNBUSINESS-Test 和 GNBUSINESS-Unlabeled 分别。一位共同作者和一位外部注释者手动将新闻集群中的事件标记为黄金标准。对于每个新闻集群，他们为每个实体分配事件中的参与者或其头部单词的预先插槽。开发集中每个插槽实现的注释器间协议 (IAA) 都有一个 Cohen 的 kappa (科恩, 1960) $\kappa = 0.7$ 。

每个数据拆分的统计信息如表所示 1，以及 与现有事件提取和事件模式归纳数据集的比较，包括 ASTRE (Nguyen 等., 2016a, MUC 4, ACE 2005² 和 ERE³，如表所示 2。与其他数据集相比，GNBUSINESS 拥有更多的文档 (即 GNBUSINESS 中的新闻集群) 以及相当数量的标签文档。

4 方法

我们研究了 ODEE 的三个递增更复杂的神经潜变量模型。

4.1 模型 1

我们的第一个模型如图所示 2(a)。它可以被视为神经延伸 Nguyen 等。

²<https://catalog.ldc.upenn.edu/>

LDC2006 6T06

³<https://catalog.ldc.upenn.edu/>
LDC2013E64

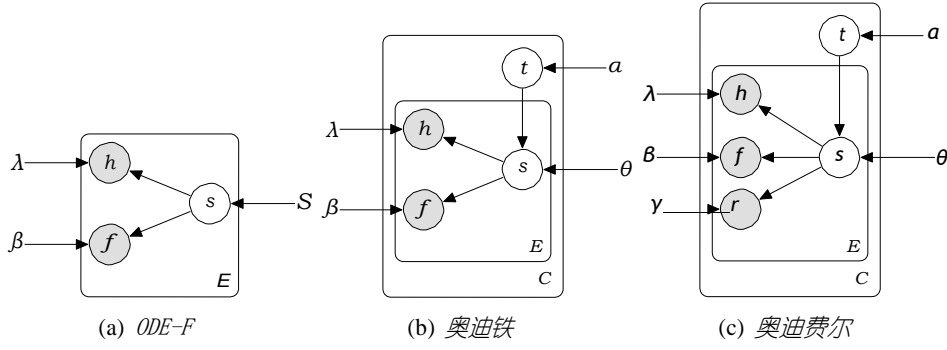


图 2: 模型的铭牌表示法。(S - 插槽数; E - 实体数; C - 新闻集群数; V - 头字词汇量; 灰色圆圈是观察到的变量, 白色圆圈是隐藏变量。)

算法 1 ODEE-F

- 1: 对于每个实体 $e \in E$ do
- 2: 采样槽 $s \sim \text{Uniform}(1, S)$
- 3: 采样头 $h \sim \text{Multinomial}(1, \lambda s)$
- 4: 采样特征向量 $f \sim \text{Normal}(\beta)$
- 5: 结束

(2015). 给定一个语料库 N , 我们从 S 个时隙的均匀分布中为每个实体 e 抽取一个时隙 s , 然后从一个多媒体中为一个头字 h 采样, 以及由上下文编码器产生的连续特征向量 $f \in \mathbb{R}^n$ 。

为简单起见, 我们假设 f 遵循多变量正态分布, 其协方差矩阵是对角矩阵。我们标记 S 个不同正态分布的所有参数 (平均向量和协方差矩阵的对角向量)。

f 为 $\beta \in \mathbb{R}^{S \times 2n}$, 其中 n 代表 f 的维数, 处理概率矩阵

槽头分布中的 $\lambda \in \mathbb{R}^{S \times V}$ 作为参数

在行式单纯形约束条件下, 其中

V 是头词词汇量大小。我们称这个型号为 ODEE-F。

预先训练的情境化嵌入, 如 ELMo (彼得斯等人., 2018), GPT (Radford 等., 2018, 2019) 和 BERT (Devlin 等人., 2018) 改进一系列自然语言

通过提供丰富的语言模型信息来处理任务。

我们选择 ELMo⁴ 作为我们的背景

tual 特征编码器, 通过使用字符表示来处理未知单词。

生成故事在算法中显示 1.

实体 e 的联合概率是

$$p_{\lambda, \beta}(e) = p(s) \times p_{\lambda}(h|s) \times p_{\beta}(f|s) \quad (1)$$

⁴ 在实践中, 我们使用 “小” ELMo 模型, 输出为 $2 \times$

算法 2 ODEE-FE

- 1: 对于每个新闻集群 $c \in N$ do
- 2: 采样潜在事件类型向量 $t \sim \text{Normal}(\alpha)$
- 3: 对于每个实体 $e \in E$ do
- 4: 采样槽 $s \sim \text{Multinomial}(\text{MLP}(t; \theta))$
- 5: 采样头 $h \sim \text{Multinomial}(1, \lambda s)$
- 6: 采样特征向量 $f \sim \text{Normal}(\beta s)$
- 7: 结束
- 8: 结束

4.2 模型 2

ODEE-F 的限制是来自全局均匀分布的采样时隙分配不足以模拟不同事件可能具有不同时间分布的事实。因此, 在图

尤里 2(b), 我们进一步从全局中为每个新闻集群抽样潜在事件类型向量 $t \in \mathbb{R}^n$

由 α 参数化的正态分布。然后, 我们使用 t 和带参数 θ 的多层感知器 (MLP) 来编码相应的时隙分布对数, 对离散时隙分配进行采样

$s \sim \text{Multinomial}(\text{MLP}(t; \theta))$ 。的输出

MLP 通过 softmax 层之前 -

使用过。我们将此模型命名为 ODEE-FE。

生成故事在算法中显示 2.

新闻集群 c 的联合概率是

$$p_{\alpha, \beta, \theta, \lambda}(c) = p_{\alpha}(t) \times p_{\theta}(s|t) \times p_{\lambda}(h|s) \times p_{\beta}(f|s) \quad (2)$$

4.3 模型 3

128-dhttps://allennlp.org/elmo 作为初始参数并在 GNBUSINESS-Full-Text 上进行微调。

直观地说，共享实体出现在新闻集群中的频率越高，它就越有可能出现重要的插槽。除此之外，不同的新闻机构关注事件论证的不同方面，它们可以通过文本冗余提供补充信息。一个直觉

算法 3 ODEE-FER

```

1: 对于每个新闻集群  $c \in \text{Ndo}$ 
2:   采样潜在事件类型向量  $t \sim \text{Normal}(\alpha)$ 
3:   对于每个实体  $e \in \text{Ecdo}$ 
4:     采样槽  $s \sim \text{多项式}(\text{MLP}(t; \theta))$ 
5:     采样头  $h \sim \text{多项式}(1, \lambda s)$ 
6:     采样特征向量  $f \sim \text{Normal}(\beta s)$ 
7:     采样冗余率  $r \sim \text{Normal}(\gamma s)$ 
8:   结束
9: 结束

```

它是出现频率是字级冗余的直接量度。因此，在图中 2(c)，我们另外引入了一个共同的归一化发生频率

作为观察到的潜在变量 r 的槽实现

普通 (γS)。我们称这个型号为 ODEE-FER。

形式上，新闻集群 c 接收潜在事件类型向量 t ，其中每个实体 $e \in \text{Ec}$ 接收 a

槽类型。生成故事在算法中显示 3。具有主题词，冗余上下文特征和潜在事件类型的新闻集群的联合分布是

$$\begin{aligned}
 P_{\alpha, \beta, \gamma, \theta, \lambda}(C) = & P_{\alpha}(T) \times P_1(S|T) \times P_2(S|T) \\
 & \times P_{\lambda}(H|S) \times P_6(F|S) \\
 & \times P_{\gamma}(R|S) \quad (3)
 \end{aligned}$$

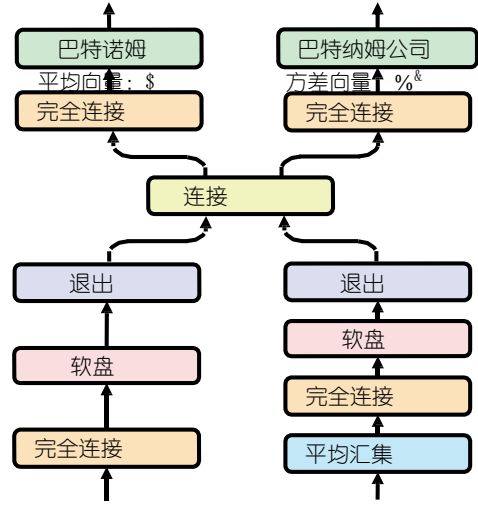
4.4 推理

我们现在考虑 ODEE-FER 的两个任务：(1) 学习参数和 (2) 在给定新闻聚类 c 的情况下进行推理以获得潜在变量 s 和 t 的后验分布。我们采用了摊销的变分推理方法 [斯里瓦斯塔瓦和萨顿 \(2017\)](#)，使用神经推理网络来学习变分参数。为简单起见，我们将 f 与 r 连接起来 ODEE-FER 和 ODER 中新观察到的特征向量 f^t 将它们参数合并为 $\beta^t \in \mathbb{R}^{S \times (2n+2)}$ 。

以下 [斯里瓦斯塔瓦和萨顿 \(2017\)](#)，我们 col- 使离散潜在变量 s 失效以获得 Evidence Lower Bound (ELBO) ([金马和威灵, 2014](#)) 对数边际可能性：

$$\begin{aligned}
 & \mathbb{E}_{\beta^t} \left[\log p_{\alpha, \beta^t, \theta, \lambda}(c) \right] \\
 & = \mathbb{E}_{\beta^t} \left[\log \left(P_{\alpha}(T) P_{\lambda}(H|T) P_{\beta^t}(F|T) P_{\gamma}(R|T) \right) \right] \\
 & = \mathbb{E}_{\beta^t} \left[\log p_{\beta^t, \theta, \lambda}(c | t) \right] - \text{DKL} \\
 & \quad [q_{\omega}(t) p_{\alpha}(t)] \quad (4)
 \end{aligned}$$

其中 $\text{DKL}[q_{\omega}||p_{\alpha}]$ 是变分后 q_{ω} 和先前 p_{α} 之间的 KL 偏差。应有



头字直方图: h

背景特征: u

图 3: 推理网络的框架。

由于难以计算不同类别的分布之间的 KL 差异以及存在简单有效的正态分布重新参数化技巧，我们选择 $q_{\omega}(t)$ 为由 ω 参数化的正态分布，这是由神经推理网络学习的。如图所示 3，我们的推理网络采用头字直方图 h (每个头字出现在新闻集群中的时间) 和上下文 t 将 f^t 作为输入，并计算平均值向量 μ 和 $q_{\omega}(t)$ 的方差向量 σ^2 。

方程 4 可以通过获得蒙特卡罗样本并应用第一学期的重新参数化技巧，并使用封闭形式的 KL 分歧项来解决。然后我们使用 ADAM 优化器 ([金马和巴, 2014](#)) 最大化 ELBO。另外，为了缓解组件崩溃问题 ([丁和迪穆兰, 2016](#))，我们遵循 [Srivastava 和 Sutskever \(2017\)](#) 并在 ADAM 优化器中使用高力矩权重 (> 0.8) 和学习率 ($[0.001, 0.1]$)，执行批量标准化 ([约费 and 塞格迪, 2015](#)) 和 dropout ([Srivastava 等 2014](#))。在学习了模型之后，我们制作了插槽

MLE 提及的每个实体的签名，选择最大化可能性的槽

$$\begin{aligned}
 p_{\beta^t, \theta, \lambda}(s|e, t) & \propto p_{\beta^t, \theta, \lambda}(s, h, f^t, t) \\
 & = p_{\theta}(s|t) \times p_{\lambda}(h|s) \times p_{\beta^t}(f^t|s) \quad (5)
 \end{aligned}$$

名称	值
插槽号 S.	30
特征维度	256
完全连接的层大小	100
MLP 层数	1
激活功能	软盘
学习率	0.002
动量	0.99
辍学率	0.2
批量大小	200

表 3: 超参数设置。

4.5 组装输出事件

要在新闻集群 c 中组合事件以进行最终输出, 我们需要为每个实体找到谓词, 该谓词现在具有槽值。我们使用由 Stanford 依赖解析器生成的 POS 标签和解析树 (克莱恩和曼宁, 2003) 提取每个实体提及的首字的谓词。应用以下规则: (1) 如果头字的管理者是 VB, 或者 (2) 如果头字的管理者是 NN 并且属于 WordNet 的 noun. ACT 或 noun. EVENT 类别, 那么它是被视为谓词。

我们将同一个共同链中的实体提及的谓词合并为谓词集。对于这些集合中的每个谓词 v , 我们找到其谓词集合包含 v 的实体, 将实体视为由 v 触发的事件的参数。最后, 通过对参数的数量进行排序, 我们获得前 N 个开放域事件作为输出 E_c 。

5 实验

我们验证神经潜在变量的有效性

能够为 ODEE 建模和冗余信息, 并进行案例分析。我们所有的实验都是在 GNBussiness 数据集上进行的。请注意, 我们不会比较我们的模型和 MUC 4 或 ACE 2005 上的现有工作, 因为这些数据集不包含新闻集群。

设置。我们的模型和推理网络中的超参数如表所示 3。大多数超参数直接跟随斯里瓦斯塔瓦和萨顿 (2017), 根据开发实验选择槽号 S。

5.1 评估指标

模式匹配。我们遵循以前的工作, 并使用精度, 召回和 F1 分数作为模式匹配的指标 (钱伯斯和 Juraf- 天空, 2011; 钱伯斯, 2013; Cheung 等., 2013; Nguyen 等., 2015; Sha 等人., 2016; 安, 2017)。

模型答案和参考之间的匹配基于首字。在之前的工作之后, 我们将实体短语中最右边的单词或者在第一个“of”, “that”, “which”和 “by” (如果有) 之前的最右边的单词视为首字。

此外, 我们还在模型学习的插槽和注释中的插槽之间执行插槽映射。继之前的 MUC 4 工作 (钱伯斯, 2013; Cheung 等., 2013; 阮等., 2015; Sha 等人., 2016; 安, 2017), 我们实现自动贪心槽映射。每个参考时隙被映射到学习的时隙, 该时隙根据 GNBussiness-Dev 上的 F1 得分度量排名最佳。

插槽连贯性。已经提出了定性主题一致性评估的若干度量。Lau 等人. (2014) 表明, 在一组主题中所有单词对之间的归一化逐点互信息 (NPMI) 最能与所有竞争指标中的人类判断相匹配。因此, 我们将其视为时隙一致性⁵。

形式上, 槽的插槽相干性 $C_{NPMI}(s)$ s 通过使用其前 N 个头部单词计算得出

$$C_{NPMI} = \frac{1}{N} \sum_{i=2}^N \sum_{j=1}^{i-1} NPMI(w_i, w_j) \quad (6)$$

$$NPMI(w_i, w_j) = \frac{p(w_i, w_j) + E}{- \log(p(w_i) p(w_j)) + E} \quad (7)$$

其中 $p(w_j)$ 和 $p(w_i, w_j)$ 是基于估计的在外部参考文档上的滑动窗口内导出的单词共现计数, 并添加 E 以避免零对数。

以前关于主题一致性的工作使用维基百科和千兆字作为参考语料库来计算单词频率 (纽曼等人., 2010; Lau 等人., 2014)。我们使用 GNBussiness-Full-Text, 其中有 1.45M 个句子和 31M 个单词, 足以估计概率。为了减少稀疏性, 对于每个新闻报道, 我们计算整个文档中的单词共现而不是滑动窗口。此外, 对于每个插槽, 我们保留前 5 个, 前 10 个, 前 20 个和前 100 个单词, 平均测试集上的 4xS 一致性结果。

⁵ 我们使用实现 <https://github.com/JHLAU/拓扑解释性>。

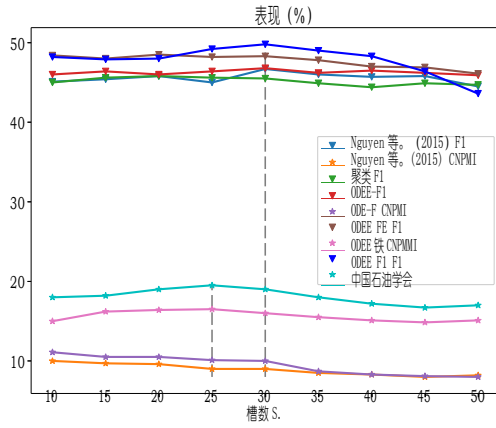


图 4: 具有不同数量的插槽 S 的五个模型的模式匹配和平均时隙相干性的 CN 分数。

方法	架构匹配 (%)		
	P	R	F ₁
Nguyen 等. (2015)	41.5	53.4	46.7
聚类	41.2	50.6	45.4
ODEE-F	41.7	53.2	46.8
奥迪铁	42.4	56.1	48.3
奥迪费尔	43.4	58.3	49.8

表 4: 模式匹配的整体性能。

5.2 开发实验

我们在 GNBussness-All 上学习模型，并使用 GNBussness-Dev 通过 [10, 50] 中的网格搜索确定插槽编号 S，步长等于 5。4 显示了我们在下一小节中介绍的五个模型的模式匹配和平均时隙相干性的 F1 得分，不同数量的时隙 S 范围从 10 到 50。我们可以看到，对于 ODEE-FER 的最佳 F1 得分，最佳数量时隙为 30，而对于最佳时隙相干性，最佳时隙数为 25。大于 30 或小于 25 的 S 值对 F1 得分和时段相干性的结果较低。考虑到 F1 得分和时段相干性之间的平衡，我们选择 $S = 30$ 作为剩余实验的最终 S 值。

5.3 最终结果

表 4 和表 5 显示最终结果。该

基于适当的 t 检验的 p 值是

方法	Ave Slot Conherence
Nguyen 等. (2015)	0.10
ODEE-F	0.10
奥迪 FE 奥	0.16
迪 FER	0.18

表 5: 平均时隙一致性结果。

在比较值接近的情况下，请参见下文。我们比较我们的工作阮 等. (2015)，MUC 4 上的现有技术模型将每个实体表示为包含头字，属性关系特征列表和谓词关系特征列表的三元组。模型中的要素是离散的，并从依赖解析树中提取。除特征外，模型结构与 ODEE-F 相同。

测试我们的外部功能的优势

孤立地，我们通过 ODEE-F 中采用每个实体的连续特征并运行谱聚类来构建另一个基线模型 (冯 卢克斯堡, 2007)。我们称之为聚类。

模式匹配。表 4 显示了 GNBussness-Test 上模式匹配的整体性能。从表中可以看出，ODEE-FER 在所有方法中都获得了最佳 F1 分数。通过比较 Nguyen 等. (2015) 和 ODEE-F ($p = 0.01$)，我们可以看到使用连续上下文特征比离散特征提供更好的性能。这证明了连续上下文特征在减轻文本中离散特征的稀疏性方面的优势。我们还可以从聚类的结果中看出，仅使用上下文特征对于 ODEE 来说是不够的，同时与我们的神经潜在结合

ODEE-F 中的变量模型可以获得强有力的结果 ($p = 6 \times 10^{-6}$)。这表明神经网络帐篷变量模型可以更好地解释观察到的数据。

这些结果证明了我们的方法在结合上下文特征，潜在事件类型和冗余信息方面的有效性。在 ODEE 模型中，ODEE-FE 对 ODEE-F 的 F1 得分增加 2%，这表明潜在事件类型建模是有益的，并且时隙分布依赖于潜在事件类型。此外，通过比较 ODEE-FER 和 ODEE-FE，F1 得分增加 1%

($p = 2 \times 10^{-6}$)，这证实了杠杆作用冗余也有利于探索哪些插槽应该分配实体。

插槽连贯性。表 5 显示了模式中所有时隙的平均时隙相干结果的比较。请注意，我们不报告群集模型的插槽一致性，因为它不会在每个插槽中输出前 N 个头字。ODEE-FER 的平均时隙相干性最高，这与表中的结论一致 4. 平均时隙一致性

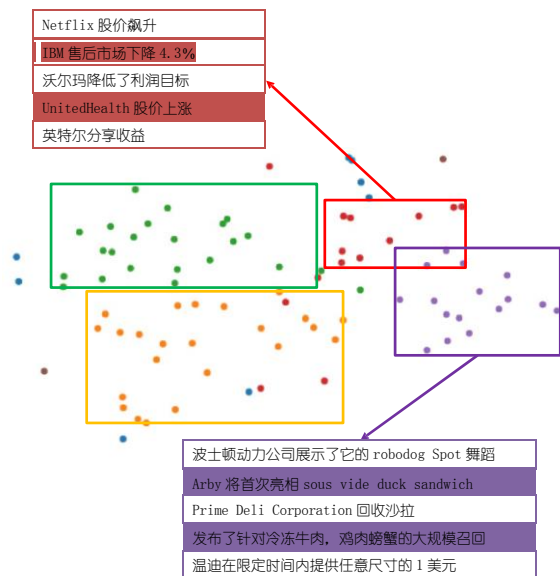


图 5: 具有通过光谱聚类产生的彩色标签的测试集中的潜在事件类型向量的 T-SNE 可视化结果。

ODEE-F 的可比性 *Nguyen 等。 (2015)* ($p = 0.3415$)，这再次证明了上下文特征是离散特征的强有力替代。ODEE-FE 的得分 ($p = 0.06$) 和 ODEE-FER ($p = 10^{-5}$) 都是高于 ODEE-F，证明潜在事件类型在 ODEE 中是关键的。

5.4 潜在事件类型分析

我们感兴趣了解潜在事件类型向量的建模程度。为此，对于 GNBUSINESS-Dev 中的每个新闻集群，我们使用图中的推理网络 3 计算潜在事件类型向量 t 的平均 μ 。T-SNE 转型 (Maaten 和 Hinton, 2008) 平均向量如图所示 5. 进一步应用光谱聚类，并通过 Calinski-Harabasz 得分选择聚类数量 (Calinski 和 Harabasz, 1974) 在网格搜索中。

在图中 5，有四个以不同颜色标记的主要簇。新闻报道的代表性标题显示为示例。我们发现向量显示每个主要群集的显着主题。例如，红色集群包含有关 Netflix 股价飙升，IBM 下跌，英特尔股票收益等股票上涨和下跌的新闻报道；紫色集群中的新闻报道主要是与产品相关的活动，例如 Boston Dynam' 揭示它的 robodog Spot 舞蹈，Arby 将首次亮相 sous vide duck sandwich，Wendy 的发行 \$ 1 任何大小的油炸等绿色群集和

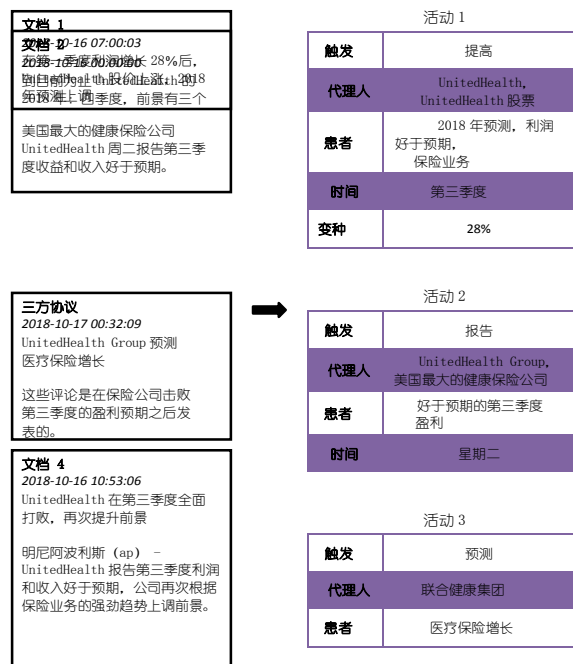


图 6: UnitedHealth 股票的提取开放域名事件上升。

橙色集群也是可以解释的。前者是关于组织报告的变化，而后者是关于服务相关的活动。

5.5 案例分析

我们进一步使用新闻集群 UnitedHealth 股票上涨图 5 用于案例研究。数字 6 显示从新闻集群中提取的前 3 个开放域事件，其中左侧显示了四个输入新闻报告，右侧显示了三个系统生成的事件，带有映射的插槽。

通过比较普通新闻报道和提取的事件，我们可以看到输出事件分别为“新闻”，“报告”和“预测”触发的三个事件提供了新闻集群的合理摘要。大多数插槽都是有意义的，并且与触发器密切相关，同时涵盖了大多数关键方面。但是，此示例还包含几个不正确的插槽。在事件 1 中，时隙“变化”及其实现“28%”仅与实体“好于预期的利润”相关，但事件中有三个时隙实现，这引起混淆。此外，插槽“Aim”没有出现在第一个事件中，其实现应该是文档 1 中的“第三季度利润”。原因可能是我们仅使用具有相同谓词的实体来组装事件，这会引入噪声。此外，由于

解决共同参考链中的预处理错误，输出中缺少某些实体提及。

还有一种情况，其中一个插槽实现在语义上与一个触发器相关，但最终出现在不同的事件中。一个例子是实体“好于预期的利润”，它与谓词“报告”有关，但最终出现在“加薪”事件中。原因可能是从解析依赖树传播的错误，这会混淆实体头字的句法谓词。

6 结论

我们提出了开放域事件提取的任务，从新闻集群中提取无约束类型的事件。研究了一种新的潜变量神经模型，该模型探讨了潜在事件类型向量和实体提及冗余。此外，GNBusiness 数据集是一个带有各种事件类型和可解释事件模式的大型数据集，随本文一起发布。据我们所知，我们是第一个使用神经潜变量模型来诱导事件模式和提取事件的人。

致谢

我们感谢匿名审稿人提出的宝贵意见和建议。我们感谢河内科技大学的 Kiem-Hieu Nguyen 为他们的工作提供源代码和解决混乱。我们感谢马萨诸塞大学阿默斯特分校的 Katherine Keith 分享了有关概率模型的宝贵经验。国家自然科学基金 61751201 号，国家重点研究发展计划 2016QY03D0602，北京市科学技术委员会研究基金 Z181100008918002，Rxhui 公司资金支持⁶和中国奖学金委员会编号 201806030142。当小刘访问岳章时，工作就完成了。

参考

娜塔莉安。2017 年。诱导事件类型和角色 反向：使用函数发现主题。在“新闻研讨会的事件和故事的论文集@ ACL 2017”，第 66-76 页。

⁶<https://rxhui.com>

Collin F. Baker, Charles J. Fillmore 和 John B. Lowe。1998 年。伯克利框架网项目。在计算语言学协会第 36 届年会论文集和第 17 届计算语言学国际会议论文集，第 86-90 页。

Hila Becker, Mor Naaman 和 Luis Gravano。2011。超越热门话题：真实世界事件识别 twitter 上的阳离子。在第五届国际网络日志和社交媒体会议论文集，第 438-441 页。

Edward Benson, Aria Haghighi 和 Regina Barzilay。2011。社交媒体提要中的事件发现。“计算语言学协会第 49 届年会论文集”，第 389-398 页。

Tadeusz Calin'ski 和 Jerzy Harabasz。1974。聚类分析的树突方法。统计学中的通信 - 理论与方法，3 (1) : 1-27。

Nathanael Chambers。2013。事件架构归纳 使用概率实体驱动模型。在 2013 年自然语言处理经验方法会议论文集，第 1797-1807 页。

Nathanael Chambers 和 Dan Jurafsky。2011。基于模板的信息提取无需模板。“计算语言学协会第 49 届年会论文集”，第 976-986 页。

陈玉波，徐立恒，刘康，曾道安，赵军。2015 年。通过 dy 进行事件提取 namic 多池卷积神经网络。在计算语言学协会第 53 届年会会议记录和第 7 届亚洲自然语言加工联合会自然语言处理联合会议上，第 167-176 页。

陈玉波，杨航，刘康，赵军，贾艳涛。2018。通过 hier 进行集体事件检测 具有门控多路的逻辑和偏置标记网络 水平关注机制。在 2018 年自然语言处理经验方法会议论文集，第 1267-1276 页。

Jackie Chi Kit Cheung, Hoifung Poon 和 Lucy Van-derwende。2013。概率框架感应。在 2013 年计算语言学协会北美分会会议记录：人类语言技术，第 837-846 页。

雅各布科恩 1960 年。协议系数 名义尺度。教育和心理测量，20 (1) : 37-46。

崔安琪，张敏，刘益群，马少平，张国。2012。发现突发事件 Twitter 中流行的主题标签。在第 21 届 ACM 国际信息与知识管理会议论文集，第 1794-1798 页。

- Jacob Devlin, Ming-Wei Chang, Kenton Lee 和 Kristina Toutanova. 2018. [BERT: 预训练用于语言的深度双向变压器 - 常设](#). CoRR, abs / 1810.04805.
- Laurent Dinh 和 Vincent Dumoulin. 2016 年 [训练神经贝叶斯网](#). 技术报告.
- 冯小成, 黄立福, 杜都宇, 恒吉, 秦兵, 刘挺. 2016 年 [一种语言-用于事件检测的独立神经网络](#). “[计算语言学协会第 54 届年会论文集](#)”, 第 66-71 页.
- Elena Filatova, Vasileios Hatzivassiloglou 和 Kathleen R. McKeown. 2006 年. [自动创建 do- 主要模板](#). 在 [计算语言学协会第 44 届年会和第 21 届计算语言学国际会议论文集](#), 第 207-214 页.
- 于红, 张建峰, 马斌, 姚建民, 周国栋, 朱巧明. 2011. [运用跨实体推理以改进事件提取](#). 在 [计算语言学协会第 49 届年会的论文中](#), 第 1127-1136 页.
- Lifu Huang, Taylor Cassidy, Xiaocheng Feng, Heng Ji, Clare R. Voss, Jiawei Han 和 Avirup Sil. 2016 年 [自由事件提取和事件模式归纳](#). 在 [计算语言学协会第 54 届年会会议录](#), 第 258-268 页.
- 黄瑞红和艾伦里洛夫. 2012. [自举事件提取分类器的训练](#). “[计算语言学协会欧洲分会第 13 次会议记录](#)”, 第 286-295 页.
- Georgiana Ifrim, Bichen Shi 和 Igor Brigadir. 2014 年 [使用积极过滤器在 Twitter 中检测事件 - 和分层推文聚类](#). 在 [SNOW 2014 数据挑战会议论文集](#)中, 与第 23 届国际万维网会议同期, 第 33-40 页.
- Sergey Ioffe 和 Christian Szegedy. 2015 年. [批量也 - 恶意化: 加速深度网络培训 减少内部协变量变化](#). 在 [第 32 届国际机器学习会议论文集](#), 第 448-456 页.
- Heng Ji 和 Ralph Grishman. 2008 年. [精炼事件 ex- 通过跨文档推理进行牵引](#). “[计算语言学协会第 46 届年会论文集](#)”, 第 254-262 页.
- Diederik P. Kingma 和 Jimmy Ba. 2014 年 [亚当: 一种随机优化的方法](#). CoRR, abs / 1412.6980.
- Diederik P Kingma 和 Max Welling. 2014 年 [汽车 - 编码变分贝叶斯](#). 在 [2014 年国际学习代表会议论文集](#)中.
- 丹克莱因和克里斯托弗 D. 曼宁. 2003. [交流-curate unlexicalized 解析](#). 在 [计算语言学协会第 41 届年会的会议记录中](#), 第 423-430 页.
- Giridhar Kumaran 和 James Allan. 2005 年. [运用新事件检测的名称和主题](#). 在 [2005 年自然语言处理经验方法会议论文集](#), 第 121-128 页.
- Jey Han Lau, David Newman 和 Timothy Baldwin. 2014 年 [机读茶叶: 自动评估主题一致性和主题模型质量](#). 在 [计算语言学协会欧洲分会第 14 次会议记录中](#), 第 530-539 页.
- Liliang Li, Aixin Sun 和 Anwitaman Datta. 2012. [Twevent: 基于段的事件检测 微博](#). 在 [第 21 届 ACM 国际信息与知识管理会议论文集](#), 第 155-164 页.
- 齐力, 恒吉, 梁黄. 2013. [联合活动 通过具有全局特征的结构化预测提取 功能](#). “[计算语言学协会第 51 届年会论文集](#)”, 第 73-82 页.
- Shasha Liao 和 Ralph Grishman. 2010A. [过滤在事件提取中引导的排名](#). 在 [第 23 届计算语言学国际会议论文集](#), 第 680-688 页.
- Shasha Liao 和 Ralph Grishman. 2010 年 b. [使用 doc- 用于改善事件的 ument 级别跨事件推理 萃取](#). “[计算语言学协会第 48 届年会论文集](#)”, 第 789-797 页.
- 刘健, 陈玉波, 刘康, 赵军. 2018A. [通过门控多语言注意事件检测 机制](#). 在 [第 32 届 AAAI 人工智能会议论文集](#), 第 4865-4872 页.
- 刘树林, 陈玉波, 何世柱, 刘康, 赵军. 2016 年 [利用 framenet 改进 自动事件检测](#). 在 [计算语言学协会第 54 届年会会议录](#), 第 2134-2143 页.
- 小刘, 罗志琛, 黄和艳. 2018B. [通过注意力共同提取多个事件 - 基于图形信息的聚合](#). 在 [2018 年自然语言处理经验方法会议论文集](#), 第 1247-1256 页.
- Ulrike von Luxburg. 2007 年. [关于谱聚类的教程的 TeringBay](#). [统计与计算](#), 17 (4) : 395-416.
- Laurens van der Maaten 和 Geoffrey Hinton. 2008 年. [使用 t-sne 可视化数据](#). [机器学习研究期刊](#), 9 (11 月) : 2579-2605.

- David McClosky , Mihai Surdeanu 和 Christopher D. Manning. 2011. [事件提取作为依赖 解析](#). 在 *计算语言学协会第 49 届年会论文集*, 第 1626-1635 页。
- Andrew James McMinn 和 Joemon M. Jose. 2015 年. [基于实时实体的 Twitter 事件检测](#). 在 *实验性 IR 的会议记录中, 多语言, 多模态和互动 - 第六届 CLEF 协会国际会议*, 第 65-77 页。
- Ashutosh Modi 和 Ivan Titov. 2014 年. [诱导神经网络的脚本知识](#). 在 *第十八次计算自然语言学习会议论文集*, 第 49-57 页。
- Sean Moran , Richard McCreddie , Craig Macdonald 和 Iadh Ounis. 2016 年. [加强第一个故事检测 使用文字嵌入](#). 在 *第 39 届国际 ACM SIGIR 信息检索研究与发展会议论文集*, 第 821-824 页。
- David Newman, Jey Han Lau, Karl Grieser 和 Timothy Baldwin. 2010. [自动评估主题 相关性](#). 在 *2010 年计算语言学协会北美分会会议记录: 人类语言技术*, 第 100-108 页。
- Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret 和 Romaric Besancón. 2015 年. [生成事件 具有实体消歧的模式归纳](#). 在 *计算语言学协会第 53 届年会会议记录和第 7 届亚洲自然语言处理联合会自然语言处理国际联合会议上*, 第 188-197 页。
- Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret 和 Romaric Besancón. 2016a. [用于打开的数据集 事件提取用英语](#). 在 *第 10 届国际语言资源和评估会议论文集*, 第 1939-1943 页。
- Thien Huu Nguyen, Kyunghyun Cho 和 Ralph Grishman. 2016B. [通过反复提取联合事件 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。
- Thien Huu Nguyen 和 Ralph Grishman. 2016 年. [国防部- 使用 convolu-进行事件检测的 skip-gram 国家神经网络](#). 在 *2016 年自然语言处理经验方法会议论文集*, 第 886-891 页。
- Nikolaos Panagiotou, Cem Akkaya, Kostas Tsiout-siouliklis, Vana Kalogeraki 和 Dimitrios Gunop-ullos. 2016 年. [使用实体检测 第一个故事 和关系](#). 在 *第 26 届国际计算语言学会议论文集*, 第 3237-3244 页。
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee 和 Luke Zettlemoyer. 2018. [深层语境化词代表 判刑](#). 在 *2018 年计算语言学协会北美分会会议记录: 人类语言技术*, 第 2227-2237 页。
- Karl Pichotta 和 Raymond J. Mooney. 2016 年. [运用 脚本的句子级 LSTM 语言模型 - 约见](#). 在 *计算语言学协会第 54 届年会论文集*中。
- 秦艳霞, 张悦, 张敏, 郑德权. 2013. [功能丰富的基于细分的新闻 在 Twitter 上进行事件检测](#). 在 *第六届国际自然语言处理联合会议论文集*, 第 302-310 页。
- 秦艳霞, 张悦, 张敏, 郑德权. 2017 年. [语义框架表示 在 Twitter 上进行事件检测](#). 在 *2017 年亚洲语言处理国际会议论文集*, 第 264-267 页。
- Long Qiu, Min-Yen Kan 和 Tat-Seng Chua. 2008 年. [在场景模板创建中建模上下文](#). 在 *第三届国际自然语言处理联合会议论文集*, 第 157-164 页。
- Alec Radford, Karthik Narasimhan, Tim Salimans 和 Ilya Sutskever. 2018. [改善语言 - 坚持生成性的预训练](#).
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei 和 Ilya Sutskever. 2019. [语言 模型是无监督的多任务学习者](#).
- Roi Reichart 和 Regina Barzilay. 2012. [多事件 由全局约束引导的提取](#). 在 *2012 年计算语言学协会北美分会会议论文集: 人类语言技术*, 第 70-79 页。
- Alan Ritter, Mausam, Oren Etzioni 和 Sam Clark. 2012. [从 twitter 开放域名事件提取](#). 在 *第 18 届 ACM SIGKDD 知识发现和数据挖掘国际会议论文集*, 第 1104-1112 页。
- Rachel Rudinger, Pushpendre Rastogi, Francis Ferraro 和 Benjamin Van Durme. 2015 年. [脚本归纳 作为语言建模](#). 在 *2015 年自然语言处理经验方法会议论文集*, 第 1681-1686 页。
- Takeshi Sakaki, Makoto Okazaki 和 Yutaka Matsuo. 2010. [地震震撼了 Twitter 用户: 实时 社交传感器检测事件](#). 在 *第 19 届万维网国际会议论文集*, 第 851-860 页。

雷莎, 苏建利, 鲍宝昌, 志芳穗。2016 年[联合学习模板和活动插槽 图式归纳](#). 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 428-434 页。

雷莎, 冯茜, 鲍宝昌, 志芳穗。2018。联合提取事件触发器和论证 - 依赖桥接 RNN 和基于张量的论证互动. 在第 32 届 AAAI 人工智能会议论文集中, 第 5916-5923 页。

Yusuke Shinyama 和 Satoshi Sekine。2006 年。[前置-使用不受限制的关系提取信息 发现](#). 在 *2006 年计算语言学协会北美分会会议记录: 人类语言技术*, 第 304-311 页。

Akash Srivastava 和 Charles Sutton。2017 年。[自动 en- 编码主题模型的变分推理](#). 在 *2017 年国际学习代表会议论文集中*。

Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky , Ilya Sutskever 和 Ruslan Salakhutdinov。2014 年[辍学: 一种预防神经的简单方法 网络过度拟合](#). *机器学习研究杂志*, 15 (1) : 1929-1958。

Beth Sundheim。1992 年。[第四个消息概述 圣人理解评估和会议](#). 在 *第 4 次消息理解会议论文集*, 第 3-21 页。

Jeroen BP Vuurens 和 Arjen P. de Vries。2016 年[第一 故事检测使用多个最近邻居](#). 在 *第 39 届国际 ACM SIGIR 信息检索研究与发展会议论文集*, 第 845-848 页。

Bishan Yang 和 Tom M. Mitchell。2016 年[联合前文档中事件和实体的牵引力 上下文](#). 在 *2016 年计算语言学协会北美分会会议论文集: 人类语言技术*, 第 289-299 页。

全元, 向仁, 何文琪, 张超, 耿新和, 黄立福, 恒吉, 林钦烨, 韩佳薇。2018。[开放式模式事件分析 大规模新闻语料库](#). 在 *第 27 届 ACM 国际信息和知识管理会议论文集*, 第 587-596 页。