

文章编号: 1003-0077(2019)05-0082-11

基于框架语义扩展训练集的有监督事件检测方法

张婧丽, 周文瑄, 洪宇, 姚建民, 周国栋, 朱巧明

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘要: 事件检测是信息抽取领域的一个重要研究方向, 目前的事件检测方法往往受限于数据稀疏、语料例句分布不平衡和歧义问题。该文研究发现框架语义知识库 FrameNet(FN)含有丰富的已标注框架的语料, 并且 FN 中定义的框架和事件检测中定义的事件具有极其相似的结构。框架由词法单元和一组框架元素组成, 可与事件中的触发词和论元形成对应关系; 而且, FN 中的许多框架实际上也能表达某些事件。因此, 该文利用这一相似性构建事件类型与框架类型的映射关系, 从而选取 FN 中合适的例句作为事件检测的扩充语料, 以此来优化事件检测性能。实验结果显示, 针对触发词识别任务和事件类型识别任务, 该文提出的框架语义辅助方法取得了较好的效果。

关键词: 事件检测; 信息抽取; 框架语义

中图分类号: TP391 **文献标识码:** A

Frame Semantics Based Training Data Expansion for Supervised Event Detecting

ZHANG Jingli, ZHOU Wenxuan, HONG Yu, YAO Jianmin, ZHOU Guodong, ZHU Qiaoming

(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: Event detection is an important research issue in the field of information extraction. The current methods of event detection generally suffer from data sparseness, imbalanced distribution and ambiguity. This paper proposes to construct the correspondence between event types and frames in FrameNet (FN), so as to get additional samples to train the supervised detection models. It is revealed that FN consists of richer examples of events which have been annotated with the tags of frame semantics. In addition, the frame defined in FN shares high similarity with that in ACE: e. g. the lexical units and a set of frame elements inherently correspond to the event triggers and arguments in the ACE corpus, and many frames in FN can represent certain types of events. Experimental results show that the proposed method performs well both in trigger identification and event type recognition.

Keywords: event detection; information extraction; frame semantics

0 引言

事件抽取旨在从文本中识别事件信息, 在问答、文本分析和知识图谱等研究中都有重要应用价值。自动内容抽取 (Automatic Content Extraction, ACE) 将事件抽取定义为四个任务: 触发词识别、事件类型识别、事件角色识别和事件元素识别。本文主要研究前两个任务: 触发词识别和事件类型识别。该领域将这两个任务并称为事件检测。具体如例 1 所示, 目标是识别出句子中的触发词“战斗”及

其事件类型 Attack。

例 1 我支持那些为自己国家进行战斗的人。

触发词: 战斗

事件类型: Attack(袭击)

目前, 事件检测逐步从传统机器学习方法转向神经网络模型, 性能得到较大提升^[1-5]。但英文语料数目少, 事件类型样例分布不平衡, 仍是这一任务中重要的研究难点。ACE 英文语料仅有 3 966 条句子级别的事件描述, 事件类型 Attack(袭击)的样例数为 1 537 条, 而事件类型 Nominate(任命)的样例只有 12 条, 样例数目少, 且分布严重不平衡。这使

收稿日期: 2018-04-08 定稿日期: 2018-06-26

基金项目: 国家自然科学基金(61672367, 61672368, 61773276); 国防部科技战略先导计划(17-ZLXD-XX-02-06-04)

得学习模型在样例稀疏的事件类型上训练不充分,难以识别相关事件。上述问题均会直接影响事件检测模型的性能。

为此,本文提出:利用框架语义结构与事件类型描述之间的相似性,以及 FrameNet^[6] (FN) 语料库中丰富的样本与确切的语义架构辅助事件检测。首先,采用框架类型识别方法获得 ACE 语料中触发词的框架类型,按事件类型与框架类型映射关系的强弱区分正负例框架,进而获得“正例”,用以扩充训练集。通过在相同模型上进行实验,验证该方法是否能缓解语料稀疏和样例分布不平衡问题。实验结果显示该方法对触发词识别和事件类型识别均有较好的效果。

本文的主要研究贡献如下:

(1) 提出利用 FrameNet 中的大量数据扩充 ACE 事件检测语料。

(2) 提出构建框架类型与事件类型映射关系的新方法。

(3) 检验并分析了利用 FN 中的例句扩充 ACE 语料训练集对事件检测性能的影响。

本文的组织如下:第 1 节介绍了相关工作;第 2 节陈述利用 FN 辅助事件检测的原因;第 3 节对本文中所使用的方法进行介绍,包括框架类型的识别方法与事件检测的语料处理方法;第 4 节介绍了本文采用的事件检测模型;第 5 节介绍实验架构及结果分析;第 6 节总结全文并对未来工作进行展望。

1 相关工作

目前,事件检测技术的研究思路主要包括两个方面,其一是以语义编码为基础表示学习和深度计算;其二是借助外部资源的训练数据获取与扩展。

从模型角度,事件检测技术逐步从传统机器学习方法转向神经网络方法。Li^[7]等基于结构化的感知机,将局部特征与全局特征进行结合,利用充分的句子特征信息进行事件抽取;Nguyen^[8]等将实体类型信息与词的位置信息作为特征,并利用词向量的形式表示,通过卷积神经网络(Convolutional Neural Network, CNN)自动学习这些特征,取得较好效果;Chen^[9]等提出动态多池化卷积神经网络模型(Dynamic Multi-Pooling Convolutional Neural Network, DM-CNN),即通过学习词向量得到词级的特征表示,再利用动态多池化层的卷积神经网络获得句子级的特征表示,从而保存句子的关键信息;

Tang^[10]等提出双向长短时记忆模型(Bidirectional Long Short Term Memory, Bi-LSTM),通过对句子中每个词的前面和后面的内容建模,能够很好地捕获句子中的长期依赖信息;Feng^[11]等提出将双向 Bi-LSTM 与 CNN 两个模型进行结合,先通过 Bi-LSTM 对句子中每个词的前向和后向内容进行语义编码,再经过一个 CNN 卷积神经网络捕获结构化的信息,这种联合的模型在事件检测任务上取得了很好的效果;Liu^[12]等则在人工神经网络(Artificial Neural Network, ANN)的基础上加入注意力机制,从而使模型对句子中对触发词识别与事件类型分类有重要影响的实体信息加深关注,从而有效捕获重要信息。

从语料角度考虑, Ji^[13]等将事件检测范围由单文档扩展到主题相近的文档,通过在局部特征基础上加入相近文档特征来改进事件检测性能;Hong^[14]等提出跨实体推理的事件抽取方法,即应用命名实体作为额外的判别特征辅助事件抽取;陈亚东^[15]等在 Li 等的 Joint 系统基础上,将框架语义知识作为特征加入其中,证明了框架语义作为特征的有效性;Liu^[16]等考虑到框架语义与事件在结构上的一致性,应用提前训练好的事件检测模型对 FrameNet 中的例句进行事件类型标注,同时采用框架语义中的相关规则进行约束,得到事件检测的扩充语料,改进事件检测性能;Chen^[17]等提出一种自动标注数据集的方法来获取大规模数据用于事件抽取,利用知识图谱(Freebase)自动找到一个事件的触发词,再利用 FrameNet 对触发词进行过滤和扩展,最后进行自动标注生成标注后的数据集,通过 DM-CNN 自动进行事件的抽取,实验证明该数据集可以达到与人工标注的数据集同等的效果。

FrameNet 是框架语义的典型资源库,包含了具有框架语义规则的英文单词的描述与框架标注。Narayanan^[18]等首次将 FN 应用于问答领域;之后,Shen^[19]等也在问答任务中引入 FN 知识库,获得了较好的性能。基于 FN 的研究还有 Padó^[20]等的篇章识别任务, Burchardt^[21]等的文本蕴含任务。对于这些任务, FN 中的框架知识均起到较好的作用。此外, Liu^[16]等首次将 FN 应用于事件抽取领域,利用 FN 中的例句扩充 ACE 语料,使事件抽取性能有了较大提升。

2 基于 FrameNet 的事件检测

针对引言中提出的问题,本文利用 FN 辅助事

件检测。主要原因有两个,下面分别进行介绍。

2.1 框架与事件的类型可比性分析

ACE 2005^[22]定义了 33 种事件类型,每种事件类型对应多个触发词,每个触发词包含一组样例。而 FN 是以框架语义学为基础形成的权威知识库,具体结构如图 1。FN 中定义了多种框架,每种框架包含一组词法单元,每个词法单元又包含多条已标注框架类型的样例。因此,框架类型与事件类型具有极其相似的结构,可形成较为对称的可比关系,一些框架甚至可以直接表达某些事件。表 1 列举了部分映射关系,如框架类型 *Fining* 与事件类型 *Fine* 均表示“罚款”含义。此外, FN 中的词法单元与事件中的触发词也具有对应关系。如例 2 中的触发词与词法单元均为“fractured”。

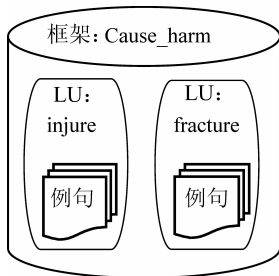


图 1 FN 中框架、词法单元与例句的关系

注:该图以框架类型 *Cause_harm*(造成伤害)为例,描述框架、词法单元与例句的关系。图 1 中的符号 LU 表示词法单元,相当于中文样例中的词。图中框架包含 38 个词法单元,这里只列举其中两类,分别为 *injure*(受伤)和 *fracture*(骨折),每个 LU 平均包含 15 个例句^①

例 2 *She has fractured legs and her right forearm.*

(译文:“她的腿和右前臂骨折了”)

触发词: *fractured*(骨折)

事件类型: *Injure*(受伤)

(注:该例句来自 ACE2005 英文事件语料)

词法单元: *fractured*(骨折)

框架类型: *Cause_harm*(造成伤害)

(注:该词法单元与框架标记来自 FrameNet 语料)

2.2 FN 含有丰富语料

FN 中含有大量词法单元,并为每个词法单元都标注了充足的例句。FrameNet 1.7 有 157 631 条例句,而 ACE 语料仅有 3 966 条。例 2 中的触发词“fractured”在 ACE 语料中只有一条例句,若作为

测试样例,那么在训练语料中将不存在用词一致的样本,从而导致两种可预见的瓶颈:

- 其一:若训练语料中不存在同义词,则即使充分的表示学习和语义编码也无法有效获取该词的语义表示,造成知识缺失,影响测试阶段的判别;
- 其二:目前,语义编码与学习的神经网络模型,在追求词一级语义共享空间的时候,需要对网络层数进行控制,当层数较少时,上述训练阶段语义缺失的样本将无法得到有效模拟,即使同义词存在于训练集,其表示的可共享性也将较低,当层数较高时,一味地追求语义表示的共性,则不仅提高计算成本(时空消耗),也将影响词义表示的独特性,从而影响模型的实用性。

但在 FN 语料中,该词含有 11 条例句,若已知框架类型 *Cause_harm* 与事件类型 *Injure* 具有映射关系,则可用这 11 条例句扩充训练集,丰富训练语料。表 1 给出了部分框架与事件类型映射表。

表 1 部分框架与事件类型映射表

框架类型	事件类型
Execution(执行刑罚)	Execution(执行刑罚)
Cause_harm(造成伤害)	Injure(受伤)
Attack(袭击)	Attack(袭击)
Fining(罚款)	Fine(罚款)
Death(死亡)	Die(死亡)

3 事件检测的语料获取

本文利用框架语义知识从 FN 中获取例句扩充训练数据,从而优化触发词识别和事件类型识别任务。这一过程包括两方面工作:其一,获取触发词的框架类型,并区分正、负框架;再通过框架标签得到触发词在 FN 中对应的例句。其二,将获得的例句加入训练集,用产生的新数据训练事件检测模型。

3.1 框架类型识别

本文采用以下两种方法进行框架类型识别。

^① <https://framenet2.icsi.berkeley.edu>

方法一：同陈亚东^[15]等、Evangelia^[23]等一样，采用框架语义分析开源工具 SEMAFOR^① 实现框架类型识别。该模型由 FN 中 157 631 条例句训练得到。本文旨在用该工具识别 ACE 语料中触发词的框架。SEMAFOR 对每个句子中的触发词能识别出唯一的框架类型，如例 3 中的框架标签 Possession(拥有)；再通过检索 FN 得到该触发词的其他框架。本文将 SEMAFOR 识别的框架指定为正例框架，该框架下的例句为“正例”，其他框架即为负例所在的框架。

例 3 *The country had banned a woman from having more than one child.*

(译文：“以前，国家禁止生二胎”)

触发词：having(有)

事件类型：Life(生命)

事件子类型：Be-Born(出生)

SEMAFOR 识别的框架：Possession(拥有)

触发词的其他框架：Giving_birth(给予生命)、Ingestion(采食)、Have_associated(有关联)、Inclusion(入选)

方法二：由于框架与事件具有较为对称的可比关系，甚至可直接用框架类型表示事件类型，所以二者的标签具有较高相似度。本文利用这一点进行框架类型识别。首先，检索 FN 得到触发词对应的所有框架类型标签，再计算这些框架标签与触发词的事件类型标签及事件子类型标签的相似度；计算框架与事件的最终相似度，从而判断得到正、负框架。具体计算方法如下：

(1) 将触发词的框架标签与事件类型标签及事件子类型标签都视作短语(即短句子)，采用 Sanjeev^[24]等计算句子词向量的方法得到每个短语的词向量表示。首先，利用式(1)计算短语中每个词的重要性 f_i ； n_i 表示第 i 个词在词表中出现的次数， a 为参数，本文设置为 10^{-3} 。再利用式(2)计算整个短语的词向量 v ； L 为短语长度，即短语中词的个数， v_i 表示第 i 个词的词向量。将每个词的词向量乘以其重要性 f_i 后求和，即对短语中的每个词按其频率分配不同重要性。再除以 L ，即得到整个短语的词向量表示。

$$f_i = \frac{a}{n_i + a} \quad (1)$$

$$v = \frac{\sum_{i=1}^n f_i v_i}{L} \quad (2)$$

(2) 计算触发词的每个框架标签与事件类型标签的相似度。 e 为该短语词向量的每个维度， t 为词向量的总维度。 e_i^a 和 e_i^b 分别表示该触发词事件类型标签与框架标签词向量的第 i 个维度，采用余弦相似度求得 s_1 ，如式(3)所示。

$$s_1 = \frac{\sum_{i=1}^t e_i^a e_i^b}{\sqrt{\sum_{i=1}^t (e_i^a)^2 \cdot (e_i^b)^2}} \quad (3)$$

(3) 按上述方法计算得到该触发词的每个框架标签与事件子类型标签的相似度(记为 s_2)。

(4) 利用式(4)，对上面得到的两个相似度结果求算术平均，得到该触发词的每个框架标签与事件标签的最终相似度 s 。

$$s = \frac{s_1 + s_2}{2} \quad (4)$$

(5) 按相似度值的大小进行分类：相似度最高的框架类型为正例框架，其余为负例框架。

此外，由于某些触发词的所有框架都不能表达事件类型，所得相似度均很低，如果仍将具有最高相似度的框架作为正例框架，用其例句作为“正例”扩充训练语料，会引入噪声，对系统造成不良影响。所以本文在实验过程中需要先设定参数 γ ，只有当 $s > \gamma$ 时，才可将其归为正例框架，否则将其归为负例框架。

为了便于与方法一进行比较，同样采用例 3 的例句对方法二进行阐述。具体如例 4。

例 4 *The country had banned a woman from having more than one child.*

(译文：“以前，国家禁止生二胎”)

触发词：having(有)

事件类型：Life(生命)

事件子类型：Be-Born(出生)

搜索 FN 得到的所有框架：Possession(拥有)、Giving_birth(给予生命)、Ingestion(采食)、Have_associated(有关联)、Inclusion(入选)

根据触发词“having”搜索 FN 语料后可得到其对应的 5 个框架。根据方法二描述，利用式(1)与式(2)得到事件类型标签、事件子类型标签与所有框架标签的词向量表示后，分别计算 5 个框架标签与事件类型标签 Life 的相似度；同样，计算得到 5 个框架标签与事件子类型标签 Be-Born 的相似度；利用

① <http://www.ark.cs.cmu.edu/SEMAFOR/>

式(4)对上述两个相似度结果进行计算,得到每个框架与事件的最终相似度值,具体结果如表 2 所示。

表 2 触发词“having”的相似度结果

框架类型	相似度
Giving_birth(给予生命)	0.37
Have_associated(有关联)	0.16
Ingestion(采食)	0.07
Inclusion(入选)	0.08
Possession(拥有)	0.15

由表 2 知,相似度最高的框架为 Giving_birth

(给予生命),所以将此框架类型作为触发词的正例框架,其余框架均为负例框架。

此外,本文列出了利用该方法获得的 ACE 所有事件的框架及例句结果,如表 3 所示。

3.2 事件触发词识别与事件类型识别语料处理

利用上述方法可获得大量“正例”,这些例句根据事件与框架的高映射关系获得。很大程度上,例句中的词法单元可作为触发词,触发对应的事件类型。本文将“正例”作为训练集扩充语料,用新产生的语料进行实验,使模型得到充分训练,从而识别出更多触发词,达到缓解语料稀疏及样例分布不平衡问题的目的。

表 3 利用 SIF 方法获得的 ACE 语料中所有事件类型及其部分可比框架

事件类型	事件子类型	部分触发词/ 词法单元	部分框架类型	原例 句数/条	扩充的 例句数/条
Conflict (冲突类)	Attack(攻击)	conflict(冲突)	Hostile_encounter(敌对)	19	2
		bomb(轰炸)	Attack(袭击)	57	5
		explosion(爆炸)	Explosion(爆炸)	22	9
	Demonstrate(示威)	protest(抗议)	Protest(抗议)	18	6
Movement (移动类)	Transport(运输)	drive(驾驶)	Operate_vehicle(操作车辆)	8	10
		smuggle(走私)	Smuggling(走私)	2	10
Contact (交往类)	Meet(会面)	talk(谈论)	Chatting(闲聊)	37	11
		visit(拜访)	Visiting(拜访)	5	6
	Phone-Write(电话、书信)	contact(联系)	Contacting(联系)	3	16
Business (商业类)	Start-Org(公司设立)	create(创建)	Cause_to_start(导致开始)	1	8
	Merge-Org(公司合并)	merge(合并)	Amalgamation(合并)	8	5
	Declare-Bankruptcy(宣告破产)	run(经营)	Operating_a_system(操纵)	1	3
	End-Org(公司倒闭)	cease(停止)	Process_stop(处理结束)	2	5
Life (生命类)	Be-Born(出生)	have(有)	Giving_birth(生育)	1	3
	Marry(结婚)	married(结婚)	Personal_relationship(个人关系)	20	12
	Divorce(离婚)	divorce(离婚)	Forming_relationships(形成关系)	19	8
	Injure(受伤)	injure(损伤)	Experience_bodily_harm(身体受到伤害)	25	8
		hurt(伤害)	Cause_harm(造成伤害)	9	7
	Die(死亡)	kill(杀害)	Killing(杀死)	166	29
		death(死亡)	Death(死亡)	77	13

续表

事件类型	事件子类型	部分触发词/ 词法单元	部分框架类型	原例 句数/条	扩充的 例句数/条
Personnel (人事类)	Start-Position(任职)	replace(接替)	Take_place_of(接替)	2	7
		hire(雇用)	Hiring(雇用)	13	10
	End-Position(离职)	fire(解雇)	Firing(解雇)	13	8
	Nominate(提名)	name(任命)	Appointing(任命)	4	5
	Elect(选举)	election(选举)	Change_of_leadership(当选)	102	4
Transaction (交易类)	Transfer-Money(钱财转移)	pay(支付)	Commerce_pay(支付)	23	18
		transfer(转让)	Transfer(转让)	1	9
	Transfer-Ownership (所有权转移)	sell(出售)	Commerce_sell(销售)	6	25
		purchase(购买)	Commerce_buy(购买)	4	28
		deal(交易)	Be_in_agreement_on_action(交易)	1	8
Justice (司法类)	Acquit(无罪释放)	acquit(开释)	Verdict(判决)	6	8
	Appeal(上诉)	appeal(上诉)	Appeal(起诉)	60	2
	Arrest-Jail(逮捕)	arrest(逮捕)	Arrest(逮捕)	40	18
		detain(拘留)	Detaining(拘留)	2	9
	Sue(起诉)	claim(宣称)	Claim_ownership(索取所有权)	3	5
	Fine(罚款)	fine(罚款)	Fining(罚款)	12	9
	Pardon(赦免)	pardon(赦免)	Pardon(赦免)	2	7
	Trial-Hearing(听证)	trial(审讯)	Trial(审讯)	66	7
	Release-Parole(释放-假释)	release(释放)	Releasing(释放)	21	3
	Charge-Indict(指控)	charge(控告)	Notification_of_charges(指控警告)	68	8
	Convict(判刑)	convict(定罪)	Verdict(判决)	44	14
	Sentence(宣判)	sentence(判刑)	Sentencing(宣判)	54	7
	Execute(执行刑罚)	execution(执行)	Execution(执行刑罚)	13	8
	Extradite(引渡)	extradition(引渡)	Extradition(引渡)	6	10

注：①由于篇幅限制，该表格对每个事件类型只列举了部分触发词及其对应的框架，且没有列举具体的例句。

② 本文所有可比的框架类型、词项和扩充语料所用的例句形成的数据集将对所有学术研究免费开放。

4 事件检测模型

为充分验证扩充语料对事件检测是否有效，本文分别采用基于特征抽取的传统机器学习模型和神经网络模型进行实验。

4.1 传统机器学习模型

基于特征抽取的传统机器学习模型采用 Li^[7]等的联合事件检测模型。所谓基于特征抽取指该模型主要通过抽取样例的各种特征实现触发词识别与

事件类型识别任务。该模型采用的具体特征如表 4 所示。主要包括词法、句法和实体信息三个层面的特征。此外，该模型采用最大熵分类器^①，同时实现触发词识别和事件类型识别，即将句子中的每个词当作候选触发词，分类器对每个候选触发词进行判断，赋予该词具体的事件类型标签。若事件类型不为空，则认为该词为触发词，其标签即为触发的事件类型；否则为非触发词，不触发事件类型。所以该模型的事件类型为 33 种，不包括 NA(空)类。

① <http://mallet.cs.umass.edu/>

表 4 触发词识别与事件类型识别特征表

特征类型	特征描述
词法	(1) 当前词左/右两个词的词性 (2) 当前词左/右两个词的词形 (3) 当前词的同义词 (4) 当前词的词形 (5) 当前词的 Brown 聚类
句法	(6) 当前词的依赖词和支配词 (7) 当前词的依存类型 (8) 当前词是否为修饰词 (9) 当前词是否为引用代词
实体信息	(10) 当前词左/右两个词的实体类型 (11) 当前词的依赖词的实体类型 (12) 句子中与当前词最近的实体类型

4.2 神经网络模型

神经网络模型方面, Liu^[25]等、Lin^[26]等采用双向长短时记忆模型(Bidirectional Long Short Time Memory, Bi-LSTM)在自然语言处理任务中均取得了较好的效果,所以本文也采用该神经网络模型进行事件检测,并严格遵循 Feng^[11]等的 Bi-LSTM 架

构。Bi-LSTM^[27]是一种双向循环神经网络,在隐藏层同时有一个正向 LSTM 和一个反向 LSTM,正向 LSTM 可以捕获当前词的上文信息,反向 LSTM 可以捕获当前词的下文信息,这样模型就可以捕获句子中每个词的完整的上下文信息。图 2 具体显示了该模型的主要架构。本文按照 Chen^[9]等的方法,将整个句子作为模型的输入。在输入前先

$$P(y_i | x_i, \theta) = \frac{e^{o_i}}{\sum_{k=1}^t e^{o_k}} \quad (5)$$

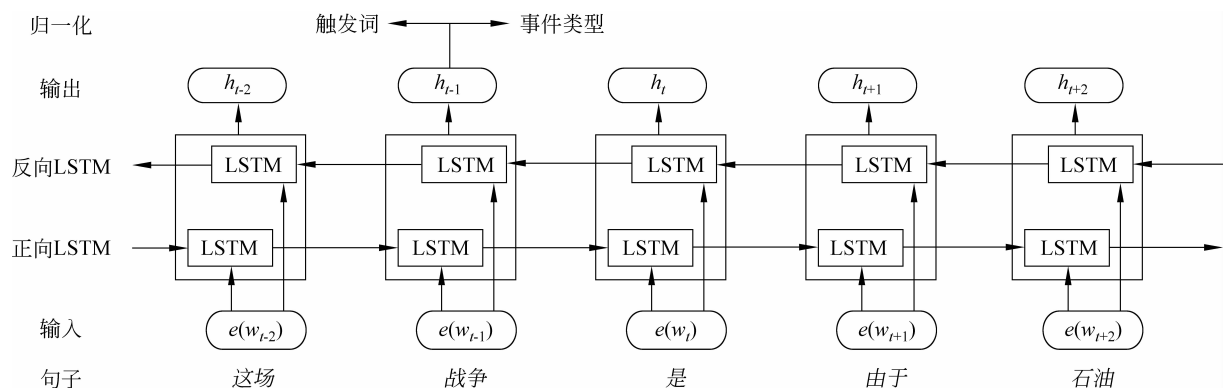


图 2 Bi-LSTM 模型架构图

该模型同样将句子中的每个词作为候选触发词,最终目标是对这些候选触发词进行分类,将其分为 34 个事件类别,包括 33 个 ACE 事件类型和一个 NA(空)类。训练过程中,采用小批量随机梯度下降^[28]方法最小化损失函数,并且添加一个 dropout 层^[29](指的是在模型训练过程中随机让网络中某些隐藏层节点的权重不工作),作用是防止出现过拟合问题。此外,该模型采用多类交叉熵作为损失函数,计算公式如下: K 为训练数据中所有的词; y_i 为词 x_i 的真实事件类型; λ 为正则化项, θ 为所有参数。

$$L(\theta) = - \sum_{i=1}^K \log P(y_i | x_i, \theta) + \lambda(\theta) \quad (6)$$

5 实验分析

5.1 数据与评估方法

针对触发词识别和事件类型识别任务,本文选取 ACE2005(Automatic Content Extraction 2005)的 599 篇英文事件文档作为实验的初始语料。为了便于对照,采用同 Li^[7]等相同的语料,选取其中的 529 篇文档为基本训练集,另外 40 篇新闻文档为测试集。扩充语料采用从 FrameNet 1.7 中获取的例句。

触发词识别任务主要考察句子中的某个词是否

触发事件类型,系统性能取决于正确识别出的触发词个数。事件类型识别任务重点考察对识别出的触发词分类是否正确。本文采用准确率 P (Precision)、召回率 R (Recall) 和 F_1 值 (F_1 -Measure) 作为评价指标。

5.2 神经网络模型中的参数设置

本实验的参数设置如下:词向量维度为 300,且通过训练得到,词向量的训练过程严格遵循 Feng^[11]等在 NYT^① 语料上利用 skip-gram 预训练词向量的方法。隐藏层大小设为 100,dropout 率设置为 0.2,最小批量即每次训练的句子数为 10,学习率设置为 0.3,L2 范式取 0,调和系数 λ 设为 10^{-3} 。上述可调整的参数均在开发集上进行调整,以寻找得到最优参数。

5.3 框架识别

5.3.1 实验系统设置

本文首先对 3.1 节提出的两种框架识别方法进行实验,从而判定采用哪种方法能获得更好的扩充语料。为此,提出以下对比实验:

- SEF: 使用 SEMAFOR 开源工具直接获取触发词的框架,按 3.1 节中介绍的第一种方法得到“正例”,并用其扩充 ACE 初始语料的训练集,测试集不变。采用第 4.1 节介绍的基于特征的事件检测模型进行触发词识别和事件类型识别。
- SIF: 使用 3.1 节介绍的第二种方法,即通过相似度计算得到框架类型与事件类型映射关系的强弱,从而区分出正、负例框架,利用得到的“正例”扩充训练集得到新的训练语料,与 SEF 采用相同的模型对事件检测任务进行实验。

- Frame_forward_combine^[15] (简称为 FRC) (2015): 利用 SEMAFOR 识别框架类型,并将其作为特征,结合 Joint 模型中的局部特征与全局特征在最大熵模型上进行实验。
- SEMAFORE test^[31] (简称为 SEF_E (2015)): 假设框架语义解析与事件抽取在结构上是一致的,故重新训练 SEMAFOR 框架语义识别开源工具来预测事件触发词。

5.3.2 实验结果与分析

上述实验结果如表 5 所示。由表可知,用 SIF 获取的扩充语料进行事件检测实验,能够得到更好的性能。相对于 FRC 与 SEF_E,事件类型识别的 F_1 值分别提高 1.0% 和 7.2%,其他指标也都有不同程度的提升。并且相对于 SEF,触发词识别和事件类型识别的 F_1 值分别提高 0.8% 和 1.9%。可见,通过计算相似度能够提高框架类型与事件类型映射的精确度,从而使扩充的语料对事件检测更有效。

表 5 通过识别框架进行事件检测的实验结果

实验系统	触发词识别 / %			事件类型识别 / %		
	P	R	F_1	P	R	F_1
FRC	74.9	63.0	68.5	72.7	60.7	66.2
SEF _E	n/a	n/a	n/a	62.6	56.8	60.0
SEF	71.4	65.7	68.4	68.1	62.7	65.3
SIF	73.6	65.2	69.2	71.5	63.4	67.2

表 6 列出采用 SEF 与 SIF 两种方法进行框架识别的部分结果。通过分析发现出现表 5 这一结果的主要原因是: SEF 方法采用的是 SEMAFOR 框架识别工具,该工具本身对框架识别的准确率较低,同时也无法识别出所有触发词的框架,如表 6 中的“protest”。

表 6 SEF 与 SIF 两种方法的部分框架识别结果

触发词	事件类型	SEF 框架	SIF 框架
fire(解雇)	End-Position(结束职位)	Use_firearm(使用枪支)	Firing(解雇)
protest(抗议)	Demonstrate(示威)	—	Protest(抗议)
call(打电话)	Phone-Write(联系)	Request(询问)	Contacting(联系)
have(有)	Be-Born(出生)	Possession(拥有)	Giving_birth(给予生命)
hit(击打)	Attack(袭击)	Impact(影响)	Attack(袭击)
drown(淹死)	Die(死亡)	Killing(杀害)	Death(死亡)

(注:单元格中的 — 表示框架为空)

① <http://mallet.cs.umass.edu/>

但利用 SIF 不但可得到所有触发词的框架类型,同时由于一些触发词的框架可以直接表示事件类型,那么这些框架标签与事件类型标签就具有较高相似度。所以,相比于使用 SEMAFOR 工具,通过相似度计算能够更加精确地确定与触发词的事件类型具有强映射关系的框架类型。如表 6 中的触发词“hit”,其事件类型为 Attack,框架类型也为 Attack,如果用 SIF 方法,通过计算,二者会得到最高的相似度值,故很容易可以确定其框架类型为 Attack。但用 SEF 方法,识别出的框架类型为 Impact,映射关系较弱。

5.4 事件检测

上述实验证明采用 SIF 方法可获得更好的扩充样例,因此,以下实验均在用这些样例扩充训练集产生的新的训练样例上进行。

5.4.1 实验系统设置

为验证本文提出的利用 FN 中的例句扩充 ACE 训练集,是否能优化事件检测性能,本文设计以下几个实验系统进行比较。

(1) 基于特征的方法

- Joint(2013)^[7]: 基于结构化感知机,并且将局部特征与全局特征进行结合。
- Cross-Entity(2011)^[14]: 将命名实体作为额外特征的推理模型。
- PSL(2016)^[32]: 捕捉事件与事件间相关性的概率软逻辑模型。

(2) 基于语料扩充的方法

- ANN-FN(2016)^[16]: 以人工神经网络为模型,利用 FrameNet 中标注的语料提高事件检测性能。
- DM-CNN+(2017)^[17]: 自动标注大规模数据,利用动态多池化卷积神经网络模型进行事件抽取。

(3) 基于神经网络的方法

- CNN(2015)^[8]: 卷积神经网络模型。
- DM-CNN(2015)^[9]: Chen 等的动态多池化卷积神经网络模型。
- Bi-LSTM(2016)^[11]: Feng 等利用双向长短时记忆模型进行事件检测。
- Joint+FN: 与 Joint 采用相同的模型进行实验,训练集采用扩充“正例”后产生的新的训练语料。
- Bi-LSTM+FN: 采用与 Bi-LSTM 相同的

模型进行实验,训练集为扩充“正例”后的训练语料。该系统用来与 Bi-LSTM 形成对比实验。

5.4.2 实验结果与分析

用“正例”扩充训练集主要是为了缓解事件检测任务遇到的语料稀疏与样例分布不平衡问题。上述几个系统的实验结果如表 7 所示。扩充“正例”后,相对于 Joint 模型,Joint+FN 的触发词识别和事件类型识别任务的 F_1 值分别提高 1.8% 和 1.3%。相对于 Bi-LSTM 模型,Bi-LSTM+FN 的事件类型识别提高 0.8%;此外,我们的方法相对于其他未进行语料处理的模型,事件类型识别的 F_1 值均有所提高;而相对于其他扩充语料的模型,也有较大优势,相对于当前通过扩充语料进行事件抽取的模型的最高性能(ANN-FN⁺),我们的方法的 F_1 值提升 0.7%,召回率也有很大提高。由此可知:事件类型与框架类型具有强映射关系,使得某些框架类型确实可以表达事件类型,这些框架下的例句在很大程度上能触发对应的事件类型。故将这些例句作为“正例”扩充训练集后,能有效起到丰富训练语料的作用。原本由于训练例句少而无法识别的某些触发词,在语料扩充后,这些触发词具有更多的训练样例,事件检测模型能得到充分训练,故能较好地对其进行识别。为此,该方法对触发词识别与事件类型识别的召回率有明显提升,极大地缓解了 ACE 语料数据稀疏的问题。正如例 2 中的触发词

表 7 触发词识别和事件类型识别的实验结果

实验系统	触发词识别/%			事件类型识别/%		
	P	R	F_1	P	R	F_1
Joint	76.2	60.5	67.4	74.5	59.1	65.9
Cross-Entity	n/a	n/a	n/a	72.9	64.3	68.3
PSL	n/a	n/a	71.7	75.3	64.4	69.4
ANN-FN ⁺	n/a	n/a	n/a	77.6	65.2	70.7
DM-CNN+ ⁺	79.7	69.6	74.3	75.7	66.0	70.5
CNN	n/a	n/a	n/a	71.8	66.4	69.0
DM-CNN	80.4	67.7	73.5	75.6	63.6	69.1
Bi-LSTM	80.1	69.4	74.3	81.6	62.3	70.6
Joint+FN	73.6	65.2	69.2	71.5	63.4	67.2
Bi-LSTM+FN	69.1	79.9	74.1	66.5	77.0	71.4

(注:单元格中的“n/a”表示该模型中未列出此任务的结果;+表示使用了额外的数据)

“fractured”(骨折),扩充“正例”后,训练语料中包含 11 条以该词为触发词的例句,训练样本充足,模型可以很好地将该词判定为触发词。

相对于 Bi-LSTM, Bi-LSTM+FN 的触发词识别与事件类型识别的召回率分别提升 10.5% 和 14.7%,效果显著。主要原因是神经网络模型对数据量较为敏感,数据量越大,模型训练越充分。扩充语料后,训练样例数目明显增多,神经网络模型可得到充分训练,能识别出更多 Bi-LSTM 无法识别的触发词,且事件类型分类也更准确,故在召回率上获得较好性能。

相对于传统机器学习模型 Joint, Joint+FN 的召回率也有较大提升,触发词识别与事件类型识别任务分别提升 4.7% 和 4.3%。主要原因除训练数据集增大,模型训练更加充分外,还由于该模型主要是基于特征抽取,更多地依赖于例句中涉及的各种词法、句法和实体信息。故扩充语料后,具有相同词法、句法或实体信息的样例数目增多,在训练过程中,模型能较全面地捕获这些特征信息,提高对例句中触发词的辨识度,正确识别出更多触发词,有效提升召回率。

但从表 7 可以看出, Joint+FN 与 Bi-LSTM+FN 相对于 baseline,虽然在召回率上均有大幅度提升,但准确率也都有所降低。相对于 Joint, Joint+FN 触发词识别与事件类型识别的准确率分别降低 2.6% 和 3.0%;相对于 Bi-LSTM, Bi-LSTM+FN 的准确率分别降低 11.0% 与 15.1%。我们分析主要原因是语料扩充中会引入部分错误样例,产生噪声。造成这一结果的原因主要有三方面:其一,虽然我们改进了构建事件类型与框架类型映射关系的方法,但仍会有部分事件类型与框架类型无法正确映射,这会导致错误的传播,使部分错误样例扩充进 ACE 语料;其二,我们提前假设与事件类型对应的框架类型下的样例能触发该事件类型,其词法单元即触发词。但这一假设并不能保证所有的样例均具有上述性质。所以,扩充的例句不能完全保证词法单元就是触发词,即使词法单元是触发词,也不一定会触发对应的事件类型,且这些样例不一定是事件句。故将据映射关系获得的例句全部作为扩充语料,也会无法避免地引入噪声。其三,在扩充语料过程中,通过触发词获取对应的框架类型,从而将该框架下的例句作为扩充语料。所以,本文默认扩充的样例中,一个例句只有一个触发词。但通过对实验输出数据的认真分析,我们发现扩充的语料并

不能保证每个例句只有一个触发词,有些例句会出现多个触发词,而本文在处理时并没有对这些词进行事件类型标记。所以,扩充的语料在类型标记上会出现错误,在模型训练过程中会造成较大影响。

由于上述三个原因造成错误样例的引入,不管对 Joint+FN 模型还是 Bi-LSTM+FN 模型,在训练过程中,这些错误样例均会对模型产生错误引导,最终造成模型对触发词识别的准确度降低,从而使系统整体的准确率降低。

6 总结

本文针对现有事件检测方法受限于数据稀疏与语料分布不平衡问题,采用框架语义知识及 FrameNet 语料对 ACE 语料训练集进行扩充。从而解决上述两个问题,实验结果显示事件触发词识别与事件类型识别任务的性能有明显提升。此外,本文提出一种通过计算框架类型标签与事件标签的相似度来识别框架的方法,通过与直接利用开源工具 SEMAFOR 获取框架的方法相比,取得了较好的性能。

但本文默认扩充的“正例”只有一个触发词,而模型在训练过程中把每个词都当作候选触发词,一旦句子中还有其他的触发词,但我们并没有对其进行标记,那么对模型的准确率会有较大影响。此外,扩充的样例并不一定是事件句,所以未来的工作重在对句子中的其他词进行处理,如果为触发词,则为其标记相应的事件类型以及对 FrameNet 中的样例进行预处理,去掉不是事件句的样例。

参考文献

- [1] David Ahn. The stages of event extraction[C]//Proceedings of the Workshop on Annotating and Reasoning about Time and Event, Sydney, Australia, 2006: 1-8.
- [2] 杨雪蓉,洪宇,马彬,等. 基于触发词和实体推理的事件关系识别方法[J]. 中文信息学报, 2014, 28(2): 100-108.
- [3] Prashant Gupta, Heng Ji. Predicting unknown time arguments based on cross-event propagation[C]//Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Suntec, Singapore, 2009: 369-372.
- [4] Thien Huu Nguyen, Kyunghyun Cho, Ralph Grish-

- man. Joint event extraction via recurrent neural networks[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 300-309.
- [5] 姜吉发. 一种事件信息抽取模式获取方法[J]. 计算机工程, 2005, 31(15): 96-98.
- [6] Charles J. Fillmore. Frame semantics and the nature of language[C]//Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech, New York, 1976: 20-32.
- [7] Qi Li, Heng Ji, Liang Huang. Joint event extraction via structured prediction with global features[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 2013: 73-82.
- [8] Thien Huu Nguyen, Ralph Grishman. Event detection and domain adaptation with convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015: 365-371.
- [9] Yubo Chen, Liheng Xu, Kang Liu, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015: 167-176.
- [10] Duyu Tang, Bing Qin, Ting Liu. Document modeling with gated recurrent neural network for sentiment classification[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 2015: 1422-1432.
- [11] Xiaocheng Feng, Lifu Huang, Duyu Tang, et al. A language-independent neural network for event detection[J]. Science China Information Sciences, 2018, 61(9): 092106.
- [12] Shulin Liu, Yubo Chen, Kang Liu, et al. Exploiting argument information to improve event detection via supervised attention mechanisms[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 2017: 1789-1798.
- [13] Heng Ji, Ralph Grishman. Refining event extraction through cross-document inference[C]//Proceedings of ACL-08: HLT, 2008: 254-262.
- [14] Yu Hong, Jianfeng Zhang and Bin Ma. Using cross-entity inference to improve event extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, USA, 2011: 1127-1136.
- [15] 陈亚东, 洪宇, 王潇斌, 等. 利用框架语义知识优化事件抽取. [J]. 中文信息学报, 2015, 29(3): 117-125.
- [16] Shulin Liu, Yubo Chen, Shizhu He, et al. Leveraging framenet to improve automatic event detection[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 2016: 2134-2143.
- [17] Yubo Chen, Shulin Liu, Xiang Zhang, et al. Automatically labeled data generation for large scale event extraction[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 2017: 409-419.
- [18] Srinu Narayanan, Sanda Harabagiu. Question answering based on semantic structures[C]//Proceedings of the 20th International Conference on Computational Linguistics. Association for Computational Linguistics, 2004: 693.
- [19] Dan Shen, Mirella Lapata. Using semantic roles to improve question answering[C]//Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning [C]//Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2007: 12-21.
- [20] Sebastian Padó and Mirella Lapata. Cross-linguistic projection of role-semantic information[C]//Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005: 859-866.
- [21] Aljoscha Burchardt, Marco Pennacchiotti, Stefan Thater, et al. Assessing the impact of frame semantics on textual entailment[J]. Natural Language Engineering, 2009, 15(4): 527-550.
- [22] George Doddington. The automatic content extraction program-tasks, data, and evaluation[C]//Proceedings of the 4th International Conference on Language Resources and Evaluation, 2004: 837-840.
- [23] Evangelia Spiliopoulou, Eduard Hovy, Teruko Mitamura T. Event detection using frame-semantic parser[C]//Proceedings of the Events and Stories in the News Workshop, Vancouver, Canada, 2017: 15-20.

(下转第 131 页)



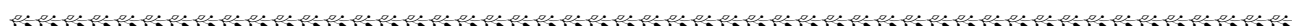
杨志明(1978—), 博士研究生, 主要研究领域为机器学习、深度学习、自然语言处理。
E-mail: yzm@ideepmind.com



王来奇(1989—), 硕士研究生, 高级工程师, 主要研究领域为机器学习、深度学习、自然语言处理。
E-mail: wlq@ideepmind.com



王泳(1975—), 通信作者, 博士后, 研究员, 主要研究领域为模式识别、自然语言处理。
E-mail: wangyong@ucas.ac.cn



(上接第 92 页)

- [24] Sanjeev Arora, Yingyu Liang, Tengyu Ma. A simple but tough-to-beat baseline for sentence embeddings [C]//Proceedings of International Conference on Learning Representation, 2016.
- [25] Pengfei Liu, Xipeng Qiu, Jifan Chen, et al. Deep fusion LSTMs for text semantic matching[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany. 2016; 1034-1043.
- [26] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, et al. A structured self-attentive sentence embedding[J]. arXiv preprint arXiv: 1703.03130, 2017.
- [27] Mike Schuster, Kuldip K Paliwal. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.
- [28] Matthew DZeiler. Adadelat: An adaptive learning rate method[J]. arXiv preprint arXiv: 1212.5701, 2012.
- [29] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [30] Tomas Mikolov, Wen-tau Yih, Geoffrey Zweig. Linguistic regularities in continuous space word representations[C]//Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013; 746-751.
- [31] Alex Judea, Michael Strube. Event extraction as frame-semantic parsing[C]//Proceedings of the 4th Joint Conference on Lexical and Computational Semantics, Denver, Colorado, 2015; 159-164.
- [32] Shulin Liu, Kang Liu, Shizhu He, et al. A probabilistic soft logic based approach to exploiting latent and global information in event classification [C]//Proceedings of the 13th AAAI Conference on Artificial Intelligence, 2016.



张婧丽(1993—), 硕士, 主要研究领域为信息抽取、自然语言处理。
E-mail: jlzhang05@gmail.com



周文瑄(1991—), 硕士, 主要研究领域为信息抽取、自然语言处理。
E-mail: chrisnotkris7@gmail.com



洪宇(1978—), 副教授, 通信作者, 主要研究领域为话题检测、信息检索和信息抽取。
E-mail: tianxianer@gmail.com