

# 基于多任务学习的中文事件抽取联合模型<sup>\*</sup>

贺瑞芳<sup>1,2</sup>, 段绍杨<sup>1,2</sup>

<sup>1</sup>(天津大学 智能与计算学部, 天津 300350)

<sup>2</sup>(天津市认知计算与应用重点实验室, 天津 300350)

通讯作者: 贺瑞芳, E-mail: rfhe@tju.edu.cn



**摘要:** 事件抽取旨在从非结构化的文本中提取人们感兴趣的信息,并以结构化的形式呈现给用户.当前,大多数中文事件抽取系统采用连续的管道模型,即:先识别事件触发词,后识别事件元素.它容易产生级联错误,且处于下游的任务无法将信息反馈至上游任务,辅助上游任务的识别.将事件抽取看作序列标注任务,构建了基于 CRF 多任务学习的中文事件抽取联合模型.针对仅基于 CRF 的事件抽取联合模型的缺陷进行了两个扩展:首先,采用分类训练策略解决联合模型中事件元素的多标签问题(即:当一个事件提及中包含多个事件时,同一个实体往往会在不同的事件中扮演不同的角色).其次,由于处于同一事件大类下的事件子类,其事件元素存在高度的相互关联性.为此,提出采用多任务学习方法对各事件子类进行互增强的联合学习,进而有效缓解分类训练后的语料稀疏问题.在 ACE 2005 中文语料上的实验证明了该方法的有效性.

**关键词:** 多任务学习;条件随机场(CRF);事件抽取

**中图法分类号:** TP18

中文引用格式: 贺瑞芳,段绍杨.基于多任务学习的中文事件抽取联合模型.软件学报,2019,30(4):1015-1030. <http://www.jos.org.cn/1000-9825/5380.htm>

英文引用格式: He RF, Duan SY. Joint Chinese event extraction based multi-task learning. Ruan Jian Xue Bao/Journal of Software, 2019,30(4):1015-1030 (in Chinese). <http://www.jos.org.cn/1000-9825/5380.htm>

## Joint Chinese Event Extraction Based Multi-task Learning

HE Rui-Fang<sup>1,2</sup>, DUAN Shao-Yang<sup>1,2</sup>

<sup>1</sup>(College of Intelligence and Computing, Tianjin University, Tianjin 300350, China)

<sup>2</sup>(Tianjin Key Laboratory of Cognitive Computing and Applications, Tianjin 300350, China)

**Abstract:** Event extraction aims to extract the interesting and structured information from unstructured text. Most Chinese event extraction methods use a continuous pipeline model which first identify event trigger word, and then identify the event arguments. Thus, it is prone to produce cascading errors, and the information contained in downstream task cannot be fed back to the upstream task. In this study, event extraction is considered as a sequence labeling task, and a multi-task learning with CRF enhanced Chinese event extraction model is proposed. Two extensions on the CRF based event extraction model are performed: (1) the separate training strategy to solve multi-label problem for an event argument in the joint model (i.e., when an event scope includes multiple events, the same entity tends to play different roles in different events); (2) considered event arguments of sub-events under the same class have the high correlation, a multi-task learning approach is proposed to jointly learn sub-events, which can alleviate the corpus sparsity to some extent. The experiment results on ACE 2005 Chinese corpus show the effectiveness of the proposed method.

**Key words:** multi-task learning; condition random field (CRF); event extraction

\* 基金项目: 国家自然科学基金(61472277); 天津市自然科学基金(18JCYBJC15500)

Foundation item: National Natural Science Foundation of China (61472277); Tianjin Natural Science Foundation of China (18JCYBJC15500)

收稿时间: 2017-03-22; 修改时间: 2017-06-02; 采用时间: 2017-08-29

人工智能的终极目标是语言理解.作为自然语言处理中信息抽取的子任务,事件抽取旨在从非结构化的文本中识别出预先定义的事件,并判断其类型及参与者.由此产生的语义理解可应用于自动文摘<sup>[1]</sup>、个性化新闻系统<sup>[2]</sup>和金融分析<sup>[3]</sup>等.相关国际评测会议——MUC(Message Understanding Conference)会议和 ACE(Automatic Content Extraction)会议的成功举办推动了该领域的技术进步并促进了工业应用的拓展.本文面向 ACE 国际评测语料(<https://catalog ldc.upenn.edu/LDC2006T06>)进行中文事件抽取的研究.

事件抽取通常被划分为两个子任务<sup>[4-9]</sup>.即:事件触发词识别/分类及事件元素识别/分类.事件触发词识别/分类的目标是从候选文本中识别出触发事件的核心词(通常为动词或者名词)并判断其类型和子类型;事件元素识别/分类的目标则是从已识别的事件中标注事件的参与者及属性(包括事件发生的时间、地点和人物等信息).例 1(文档编号: CBS20001216.1000.0355)是一个标准的 ACE 事件标注样例,它包含一个由触发词“设立”触发的 Start-Org 事件,该事件涉及两个事件元素,分别为 Place 事件元素(“新加坡”)和 Org 事件元素(“价值 36 亿美元的精元厂”).

当前,主要存在两种事件抽取模型:(i) 管道模型<sup>[4-9]</sup>;(ii) 联合模型<sup>[10,11]</sup>.前者首先完成事件触发词识别/分类,然后进行事件元素识别/分类.后者同时完成事件触发词识别/分类及事件元素识别/分类.管道模型不仅容易产生级联错误,而且处于下游的任务无法将信息反馈至上游任务,辅助上游任务的识别.例如:例 1 和例 2(文档编号: XIN20001207.0800.0071)都包含“设立”一词,该词通常作为 Start-Org 事件的触发词.对于管道模型,由于其在事件触发词识别/分类阶段无法获取到事件元素的标注信息,不仅使得模型将例 1 中的“设立”识别为 Start-Org 事件的可信值不高,还导致模型容易将例 2 中的“设立”错误地识别为 Start-Org 事件.而在联合模型中,由于让事件触发词识别/分类和事件元素识别/分类同时进行.不仅能够解决错误级联问题,而且能够有效捕捉事件触发词和事件元素的相互依赖关系.当联合模型在识别候选事件触发词“设立”时,能够捕捉到例 1 中包含 Org 事件元素(“价值 36 亿美元的精元厂”)的信息(Org 事件元素通常出现在 Start-Org 事件中).因而联合模型具有很高的可信值,将例 1 中的“设立”识别为 Start-Org 事件;而在例 2 中,由于联合模型捕捉到候选事件提及中并没有包含任何事件元素的信息,因而联合模型也具有具有很高的可信值,将例 2 中的“设立”标注为非事件(这里的非事件是指不属于 ACE 会议定义的 33 类事件).

例 1:新加坡(A1)将设立(E1)价值 36 亿美元的精元厂(A2).

例 2:咨询室将建立学生心理健康档案,设立心理信箱,开通心理咨询热线.

例 3:来自印度控制的克什米尔地区的两名武装分子(A1),前晚(A2)对印度的著名历史古迹德里红堡(A3)进行了袭击(E1),打(E2)死(E3)三人(A4).

尽管目前已有一些事件抽取联合模型的工作<sup>[10,11]</sup>,但却很少有工作关注联合模型带来的事件元素的多标签问题(当一个事件提及中包含多个事件时,同一实体往往会在不同的事件中扮演不同的角色).例如:在例 3(文档编号: ZBN20001224.0400.0009)中共包含 3 个事件: E1、E2 和 E3,触发词分别为“袭击”“打”和“死”.其中, E1 和 E2 为 Attack 事件, E3 为 Die 事件;共包含 4 个事件元素: A1、A2、A3 和 A4.其中, A1、A2、A4 是多个事件共享的事件元素.事件元素 A1 在事件提及 E1 和 E2 中扮演 Attacker 角色,而在事件提及 E3 中却扮演 Agent 角色.对于联合模型,当其遇到事件元素 A1 时,由于事件元素 A1 本身具有两个“身份”,导致联合模型无法确定应该将其标注为哪个“身份”,给模型识别带来干扰.同样的情况也发生在事件元素 A3 和 A4 中,事件元素 A3 在事件提及 E1 中扮演 Target 角色,而在事件提及 E3 中却扮演 Place 角色;事件元素 A4 在事件提及 E2 中扮演 Target 角色,而在事件提及 E3 中却扮演 Victim 角色.表 1 给出了例 3 中所有候选事件的事件触发词标注及事件元素标注的结果.

Table 1 The labeling results about Example 3

表 1 例 3 的标注结果

Event mention	Trigger	A1	A2	A3	A4
E1 (Attack)	袭击	Attacker	Time	Target	-
E2 (Attack)	打	Attacker	Time	-	Target
E3 (Die)	死	Agent	Time	Place	Victim

本文通过对 ACE 2005 中文语料的统计发现,语料中约有 36.5%的事件存在 1 个事件提及中包含多个事件的情况.为此,本文采用分类训练策略(即为每类事件分别训练一个中文事件抽取联合模型),以解决联合模型中事件元素的多标签问题.经过分类训练后,当例 3 出现在 Attack 事件对应的联合模型中时,事件元素 A1、A3 和 A4 将被分别标注为 Attacker、Target 和 Target 角色;当例 3 出现在 Die 事件对应的联合模型中时,事件元素 A1、A3 和 A4 将被分别标注为 Agent、Place 和 Victim 角色,从而有效地避免了事件元素的多标签情况.

采用分类训练策略尽管能够解决事件元素的多标签问题,但也带来了严重的数据稀疏问题.图 1 是 ACE 2005 中文语料中各类事件的分布情况.其中,Transport 事件、Attack 事件和 Die 事件的数目最多,约占事件总数的 47%;而 Acquit 事件、Extradite 事件和 Execute 事件的数目最少,约占事件总数的不到 1%.对于事件数目较少的事件,由于训练集所包含的信息量非常少,使得模型在事件数目较少的事件类别中无法获取到足够丰富的信息.因此,如何挖掘相关类别事件的关联信息变得尤为重要.通过观察发现,尽管不同类型的事件所包含的事件元素不尽相同,但处于同一事件大类下的事件子类(相关概念见第 2.1 节),其包含的事件元素有很强的关联性.如例 4(文档编号:XIN20001017.2000.0178)和例 5(文档编号:CTV20001123.1330.1541).

例 4:他说,以军(A1)还无端向拉法海关出口处的巴边防哨卡(A2)开枪,打死(E1)一名巴边防警察(A3).

例 5:另外有一名警察(A1)在交火中受伤(E1).

例 4 和例 5 分别包含了一个 Die 事件和一个 Injure 事件,触发词分别为“死”和“受伤”.虽然 Die 事件和 Injure 事件是两类不同的事件,但二者同属于 Life 事件大类.在上述两个事件提及中,事件参与者都是“警察”,并且“警察”在上述两个事件中都作为 Victim 事件元素.一个直觉的想法是能否通过这种事件子类别的相互关联关系在一定程度上弥补语料规模小的问题.为此,本文将采用多任务学习方法捕捉事件子类别之间的相互关联关系.

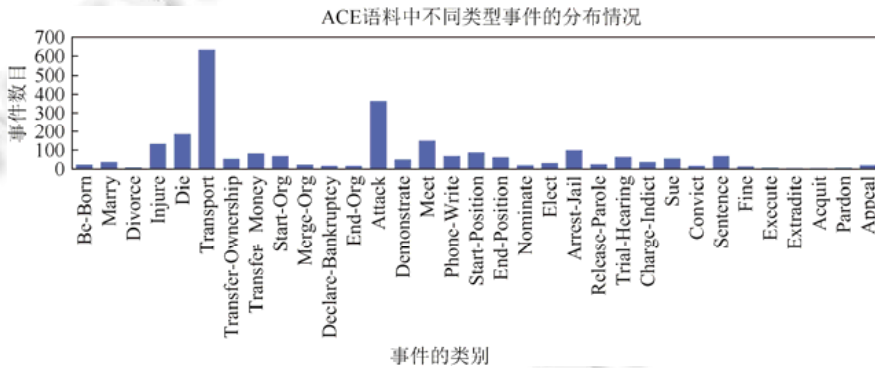


Fig.1 Event distribution in ACE 2005 Chinese corpus

图 1 ACE 2005 中文语料中各类事件的分布

本文第 1 节介绍当前中英文事件抽取方法的研究现状,并总结前人方法所存在的问题.第 2 节描述基于 CRF 的事件抽取联合模型.第 3 节描述基于 CRF 的多任务学习事件抽取联合模型.第 4 节给出实验结果与讨论.第 5 节进行总结和展望.

## 1 相关工作

目前,事件抽取研究大致分为 3 类:(1) 面向 ACE 的事件抽取<sup>[4-23]</sup>;(2) 面向生物信息的事件抽取<sup>[24-26]</sup>;(3) 开放域事件抽取<sup>[27,28]</sup>.本文工作聚焦于 ACE 事件抽取,因此下面将围绕该相关工作进行总结分析.

### 1.1 ACE 事件概述

ACE 评测会议由美国国家标准技术研究所(NIST)组织,旨在开发自动信息抽取技术以支持语言文本的自动处理,迄今为止,共举办过 8 次.在 ACE 评测会议中,事件被分为 8 个大类,33 个子类,具体见表 2.ACE 会议不仅对事件的类别进行了定义,还根据每类事件的特点定义了相应的事件元素,包括事件参与者和事件属性.下面

将介绍有关的定义及术语.

- 事件提及:包含事件的句子或者短语;
- 事件触发词:触发事件的关键词,通常是动词或者名词;
- 事件元素:参与事件的一些实体或者短语,包括事件参与者和事件属性;
- 事件类别:事件的类别由触发词的类别决定;
- 事件触发词识别/分类:识别触发事件的关键词,并判定其所属的类别;
- 事件元素识别/分类:识别参与事件的实体或者短语,并判断其在事件中扮演的“角色”.

**Table 2** ACE event type and subtype

表 2 ACE 事件类型及子类型

Type	Subtype
Life	Be-Born, Marry, Divorce, Injure, Die
Movement	Transport
Transaction	Transfer-Ownership, Transfer-Money
Business	Start-Org, Merge-Org, Declare-Bankruptcy, End-Org
Conflict	Attack, Demonstrate
Contact	Meet, Phone-Write
Personnel	Start-Position, End-Position, Nominate, Elect
Justice	Arrest-Jail, Release-Parole, Trial-Hearing, Charge-Indict, Sue, Convict, Sentence, Fine, Execute, Extradite, Acquit, Appeal, Pardon

## 1.2 英文事件抽取

现有的事件抽取研究大多聚焦于英文.其中,基于有监督的机器学习模型占主导地位,该类方法进一步分为基于特征(feature-based)的模型和基于表示(representation-based)的模型.

(1) 基于特征的模型:早期的事件抽取方法大多采用基于特征的模型,其采用 one-hot 向量表示预处理过程中提取出的词法、句法和语义特征,并将这些特征作为抽取事件的线索送入分类器中.根据选取特征范围的不同又分为句子级事件抽取模型<sup>[4]</sup>和跨句子级事件抽取模型<sup>[5-8]</sup>.前者无法处理句内信息不足或者句内信息缺失的情况.后者致力于捕捉句子之外的信息,通过更高层次的信息辅助句子级的事件抽取.代表工作有 Ji 等人<sup>[5]</sup>的跨文档推理方法;Liao 等人<sup>[6]</sup>的跨事件推理方法;Hong 等人<sup>[7]</sup>的跨实体推理方法;以及 Liao 等人<sup>[8]</sup>结合文档主题特征的事件抽取模型.

(2) 基于表示的模型:近年来,大量事件抽取研究者将目光投向基于表示的模型.在基于表示的事件抽取模型中,候选事件采用词嵌入(word embedding)表示,通过神经网络对其降维,得到其抽象表示并最终完成事件分类.相比于基于特征的模型,(i) 其无需手工提取特征,降低了人力花销;(ii) 词嵌入的引入不仅有利于模型获取语料之外的先验信息,而且使模型具有很强的泛化能力.但是由于模型需要优化大量参数,导致其需要大量标注语料以及巨额的计算开销.代表工作有 Chen 等人<sup>[12]</sup>、Nguyen 等人<sup>[13]</sup>和 Feng 等人<sup>[14]</sup>构建的基于 CNN 的事件抽取模型以及 Nguyen 等人<sup>[11]</sup>构建的基于 RNN 的事件抽取模型.

## 1.3 中文事件抽取

当前,针对中文的事件抽取研究相对较少.大致分为语言特殊性研究和基础模型研究.

(1) 语言特殊性研究:该类研究针对汉语语言的特点制定相应的事件抽取规则.其中,Chen 等人<sup>[16]</sup>聚焦中文分词与事件触发词分割的不一致性,采用全局触发词勘误表和基于字的事件触发词识别模型解决该问题并分别探索词法、句法和语义信息对中文事件抽取性能的影响;Li 等人<sup>[17]</sup>考虑汉语中词的语义通常由组成字的字决定以及中文是一种篇章驱动(discourse-driven)的语言,因而其运用组合语义和篇章一致性(discourse-consistency)识别中文未登录的事件触发词(仅出现在测试语料中的触发词).由于单纯地运用组合语义会引入大量的负样例,因而影响了模型的准确率.Li 等人<sup>[18]</sup>将组合语义、形态结构和义原相结合,有效地过滤假的事件触发词,大幅提升了模型准确率.此外,针对中文事件触发词抽取中存在的事件论元语义信息难以获取以及部分贫信息事件实例难以抽取的问题,Li 等人<sup>[20]</sup>提出了基于语义的中文事件触发词抽取联合模型.

(2) 基础模型研究:该类研究聚焦于算法的选择及建模过程的改进.Zhao 等人<sup>[21]</sup>受 Ahn<sup>[4]</sup>的启发,将事件抽取看作是一个分步进行的词分类任务.与 Ahn 不同的是,Zhao 等人在模型中加入了汉语同义词扩展,并将 Ahn 的多分类模型转化为二分类模型,有效地平衡了模型中正负样例的比例.Fu 等人<sup>[22]</sup>认为,不同的特征对模型的贡献程度不同,提出一种特征加权方法进行特征权重的再分配.然而,上述方法构建的都是管道模型.不仅容易产生级联错误,而且下游任务无法将信息反馈至上游任务,辅助上游任务的识别.因此,其后,大量研究者将工作重心转向构建中文事件抽取联合模型上.Li 等人<sup>[19]</sup>构建了事件触发词识别和事件触发词分类联合模型,并在预处理阶段运用大量推理规则过滤假的事件触发词;Chen 等人<sup>[23]</sup>在 Li 等人工作<sup>[17]</sup>的基础上分别进行了机器学习扩展和语言学扩展,并分别构建了事件触发词识别和事件触发词分类联合模型以及事件元素识别和事件元素分类联合模型;Li 等人<sup>[29]</sup>聚焦构建事件元素识别和事件元素分类联合模型,并获取句子之外的篇章级 (discourse-level) 信息,使得模型的事件元素识别/分类的  $F1$  值有较大幅度的提升.

#### 1.4 多任务学习

多任务学习是同时学习多个相关任务的机器学习算法,使用共同学习 (joint learning) 取代原先常见的独立学习 (independent learning)<sup>[30]</sup>.其不仅能利用相关任务的相互关联性缓解语料由于规模小带来的训练欠拟合问题,而且还能提升模型的泛化能力.近年来,多任务学习已成为自然语言处理的一个研究热点<sup>[31-36]</sup>,如 Liu 等人<sup>[32]</sup>采用基于多任务学习的表示学习完成语义分类及信息检索任务;Daumé 等人<sup>[35]</sup>将多任务学习模型应用于跨领域的情感分析;Prettenhofer 等人<sup>[36]</sup>通过多任务学习模型进行跨语言的文本分类等.目前在事件抽取领域只有 Liu 等人<sup>[15]</sup>有类似的工作.他们利用 FrameNet 语料辅助完成 ACE 事件抽取任务的前半部分 (事件触发词识别与事件触发词分类),但是,由于 Liu 等人没有采用机器学习算法学习两种语料的关联性,而是基于先验知识人为地构建规则来捕捉二者的相互关系,因而并不算严格意义上的多任务学习模型.考虑对 ACE 语料进行分类训练后会大幅降低训练语料规模,并且不同类型事件之间缺乏信息交互,因此,本文工作借助 Sun 等人<sup>[37]</sup>基于 CRF 的多任务学习思想,通过多任务学习方法挖掘同一大类下不同事件子类之间的相互关联关系.

综上所述,尽管目前已有一些工作聚焦于构建中文事件抽取联合模型,但本文的联合模型与这些工作不同.多数中文事件抽取联合抽取模型要么聚焦于构建事件触发词识别/分类的联合模型,要么聚焦于构建事件元素识别/分类的联合模型.很少有工作构建上述 4 个子过程的联合模型.目前, Li 等人<sup>[10]</sup>和 Nguyen 等人<sup>[11]</sup>构建了上述 4 个子过程的联合模型,但本文工作与他们的工作有所不同.本文聚焦于解决联合模型带来的事件元素多标签问题,并在联合模型之上构建了基于多任务学习的事件抽取联合模型.

## 2 基于 CRF 的事件抽取联合模型

由于管道模型不仅容易产生错误级联问题,而且无法捕捉事件触发词识别/分类对于事件元素识别/分类的后向依赖关系,为此,本文将事件抽取任务看作是序列标注任务,并采用条件随机场 (CRF) 进行建模,实现同时完成事件触发词识别/分类和事件元素识别/分类的目标.然而,将所有类别的事件混合进行训练会给模型带来事件元素多标签问题,如例 3.为此,本文将模型按照事件的类别进行分类训练,下面将给出本文构建的中文事件抽取联合模型的细节.

### 2.1 条件随机场模型

CRF 是 Lafferty 等人<sup>[38]</sup>于 2001 年提出的一种判别式的概率无向图模型.其不仅能获取上下文信息,克服 HMM 的独立性假设,还能克服 MEMMs 的标记偏置问题.其在中文分词<sup>[39]</sup>、中文命名实体识别<sup>[40]</sup>等自然语言处理任务中都有不错的效果.由于本文工作将 CRF 用于序列标注模型,因此,后文仅介绍链式 CRF.假设一个特征函数将对成的观测序列  $x$  和标注序列  $y$  映射到一个全局特征向量  $f$  中,则条件概率  $P(y|x)$  能够形式化为公式 (1)<sup>[37]</sup>.

$$p(y|x, w) = \frac{\exp[w^T f(y, x)]}{\sum_{y'} \exp[w^T f(y', x)]} \quad (1)$$

在公式(1)中, $x$  表示输入序列, $y$  表示对应输入序列  $x$  的标注序列, $f(y,x)$ 表示特征函数, $w^T$  表示特征权重。 $\sum_{y_i} \exp[w^T f(y',x)]$  是归一化项.假设训练集是包含  $n$  个样例的标注序列  $\{(x_i,y_i),i=1,\dots,n\}$ .在参数估计阶段,模型通过最大化公式(2)来求解权重参数  $w$ .

$$L(w) = \sum_{i=1}^n \log P(y_i | x_i, w) - R(w) \quad (2)$$

在公式(2)中, $\log P(y_i|x_i,w)$ 为公式(1)的对数形式; $R(w)$ 是防止过拟合的正则化项.本文采用  $L_2$  正则化,即: $R(w) = \frac{\|w\|^2}{2\sigma^2}$ , 其中, $\sigma$ 为正则化参数.为了后文表示简单,本文用  $l(i,w)$ 表示  $\log P(y_i|x_i,w)$ ,将公式(2)重写为公式(3).

$$L(w) = \sum_{i=1}^n l(i, w) - \frac{\|w\|^2}{2\sigma^2} \quad (3)$$

## 2.2 使用 CRF 构建中文事件抽取联合模型

为了联合事件触发词识别任务、事件触发词分类任务、事件元素识别任务和事件元素分类任务,本文首先借鉴了 Chen 等人<sup>[23]</sup>的思想.Chen 等人将事件触发词识别任务和事件触发词分类任务合并为一个 34 类的多分类任务(33 类事件触发词和 1 个 NULL 类(NULL 类表示该词既不是触发词也不是事件元素)),将事件元素识别任务和事件元素分类任务合并为一个 36 类的多分类任务(35 类事件元素和 1 个 NULL 类).与 Chen 等人不同的是,本文将合并后的两个子任务(34 类多分类任务和 36 类多分类任务)进一步合并为 69 类的序列标注任务(33 类事件触发词+35 类事件元素+1 个 NULL 类),并通过 CRF 模型模拟事件触发词标签和事件元素标签之间的相互依赖关系.

本文的中文事件抽取联合模型形式化为:在给定输入句子  $x=\{x_1,x_2,\dots,x_n\}$  的条件下,预测概率最大的标注序列  $y=\{y_1,y_2,\dots,y_n\}$ .其中,下标  $n$  表示句子长度; $x_i$  表示输入句子的第  $i$  个词, $x_i \in R, R$  表示中文词表; $y_i$  表示对词  $x_i$  的标注结果, $y_i \in \{Triggers \cup Arguments \cup Null\}$ .其中, $Triggers$  表示事件触发词标签(共包含 33 类标签), $Arguments$  表示事件元素标签(共包含 35 类标签).图 2 所示为本文的中文事件抽取联合模型输出样例(见例 1).

$x = \{x_1 : \text{新加坡}, x_2 : \text{将}, x_3 : \text{设立}, x_4 : \text{价值}, x_5 : \text{36亿}, x_6 : \text{美元}, x_7 : \text{的}, x_8 : \text{精元厂}\}$   
 $y = \{y_1 : \text{Place}, y_2 : \text{Null}, y_3 : \text{Start - Org}, y_4 : \text{Org}, y_5 : \text{Org}, y_6 : \text{Org}, y_7 : \text{Org}, y_8 : \text{Org}\}$

Fig.2 The labeling result of Example 1

图 2 例 1 的标注结果

## 2.3 事件元素多标签问题的解决

尽管上述模型能够让事件触发词识别/分类和事件元素识别/分类同时进行,但却无法处理同一事件元素的多标签问题.如果在一句话中包含多个事件,那么词  $x_i$  在不同的事件中往往扮演不同的角色.图 3 所示为本文的中文事件抽取联合模型对例 3 的标注结果.为了表示简单,我们只截取其中一部分,旨在举例说明事件元素的多标签问题.在图 3 中,由于  $x_7$  在不同的事件中扮演不同的角色(Victim 和 Target),联合模型将无法给  $x_7$  标注相应的标签.

$x = \{x_1 : \text{进行}, x_2 : \text{了}, x_3 : \text{袭击}, x_4 : \text{,}, x_5 : \text{打}, x_6 : \text{死}, x_7 : \text{三人}\}$   
 $y = \{y_1 : \text{Null}, y_2 : \text{Null}, y_3 : \text{Attack}, y_4 : \text{Null}, y_5 : \text{Attack}, y_6 : \text{Die}, y_7 : \text{Victim/Target}\}$

Fig.3 The labeling result of Example 3

图 3 例 3 的标注结果

为解决事件元素的多标签问题,本文将事件进行分类训练.即为每类事件分别训练一个基于 CRF 的中文事件抽取联合模型.经过预处理后的原始文本将被分别放入这些模型中,在特定类型的联合标注模型中,该联合模型仅仅标注与该类事件相对应的事件元素,因而有效地避免了事件元素的多标签问题.图 3 所示的样例将被分别送到 33 个分类器中,当图 3 所示的样例出现在 Attack 事件对应的分类器中时,将被标注为  $y'$ ,当图 3 所示的样

例出现在 Die 事件对应的分类器中时,将被标注为  $y''$ .具体如图 4 所示.

$$x = \{x_1 : \text{进行}, x_2 : \text{了}, x_3 : \text{袭击}, x_4 : \text{,}, x_5 : \text{打}, x_6 : \text{死}, x_7 : \text{三人}\}$$

$$y' = \{y_1 : \text{Null}, y_2 : \text{Null}, y_3 : \text{Attack}, y_4 : \text{Null}, y_5 : \text{Attack}, y_6 : \text{Die}, y_7 : \text{Target}\}$$

$$y'' = \{y_1 : \text{Null}, y_2 : \text{Null}, y_3 : \text{Attack}, y_4 : \text{Null}, y_5 : \text{Attack}, y_6 : \text{Die}, y_7 : \text{Victim}\}$$

Fig.4 The labeling result of Example 3 after classification training  
图 4 例 3 分类训练后的标注结果

### 3 基于 CRF 的多任务事件抽取联合模型

#### 3.1 事件的多任务划分

在基于 CRF 的事件抽取联合模型中,采用分类训练策略避免了事件元素的多标签问题,但也使得模型的训练语料规模大幅降低.因此,很难得到相对完备的训练模型.通过观察发现,在 ACE 定义的 33 类事件中,处于同一事件大类下的事件子类,其事件元素有着高度的相互关联性,如例 4 和例 5.表 3 将 Personnel 事件大类下各子类事件的事件元素进行了对比.观察表 3 可以发现:尽管事件的子类别有所不同,但其事件元素角色却极为相似.为此,本文将同一事件大类下的事件子类作为可以相互促进的多任务同时进行标注,通过多任务学习模型挖掘不同事件子类别之间的相互关联关系,进而强化事件抽取联合模型.本文共构建了 7 个基于 CRF 多任务学习的事件抽取联合模型(Movement 事件大类只包含 1 个子类,因而无法构建多任务学习模型).事件的多任务划分如图 5 所示.

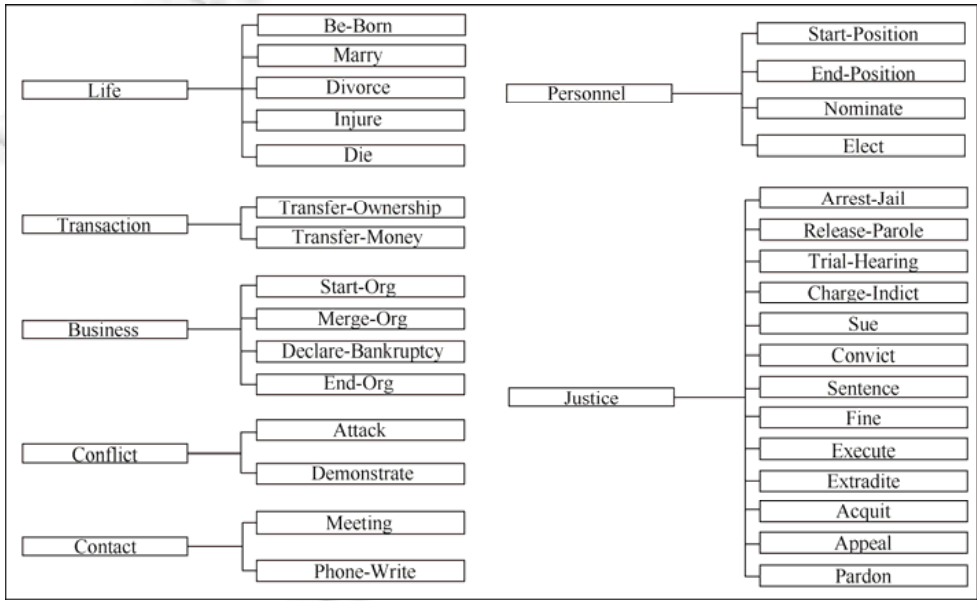


Fig.5 Multi-task division of events  
图 5 事件的多任务划分

Table 3 The comparison of event argument in Personnel category

表 3 Personnel 事件大类下各子类事件的事件元素对比

Type	Subtype	Event argument				
Personnel	Start-position	Person	Entity	Position	Time	Place
	End-position	Person	Entity	Position	Time	Place
	Nominate	Person	Agent	Position	Time	Place
	Elect	Person	Entity	Position	Time	Place

### 3.2 基于 CRF 的多任务学习模型

本节将介绍我们采用的多任务学习框架<sup>[37]</sup>.对于每个正整数  $q$ ,我们定义  $N_q=\{1,\dots,q\}$ .令  $T$  表示想要同时学习的任务数(同一事件大类下的事件子类数).对于每个任务  $t \in N_T$ ,存在  $n$  个数据样例  $\{(x_{t,i},y_{t,i}):i \in N_n\}$ ,其中  $x_{t,i}$  表示分词后的句子及其相应的特征, $y_{t,i}$  表示对词的标注.事实上,每个任务的样例数量是不同的,但是为了表示简单,我们将样例数设为定值.使用  $D$  表示  $n \times T$  的矩阵,其第  $t$  列是由数据样例  $d_t$  构成的.我们的目标能够转化为从数据  $D$  中学习特征权值矩阵  $W=\{w_1,\dots,w_T\}$ .通过最大化目标函数(见公式(4))求解  $W$ .

$$Obj(W, D) \triangleq Likelihood(W, D) - R(W) \quad (4)$$

其中,  $Likelihood(W, D)$  是所有任务的累计概率之和,见公式(5);  $R(w)$  为防止过拟合的正则化项,本文采用  $L_2$  正则化,见公式(6).

$$Likelihood(W, D) = \sum_{t \in N_T} l(w_t, D) \quad (5)$$

$$R(w) = \sum_{t \in N_T} \frac{\|w_t\|^2}{2\sigma_t^2} \quad (6)$$

公式(5)中  $l(w_t, D)$  定义为公式(7).在公式(6)中,  $\sigma_t$  为第  $t$  个任务的正则化因子;  $w_t$  为第  $t$  个任务的特征权值矩阵.

$$l(w_t, D) \triangleq \sum_{t' \in N_T} [\alpha_{t,t'} l(w_t, d_{t'})] \quad (7)$$

在公式(7)中,  $\alpha_{t,t'}$  是衡量任务  $t$  与任务  $t'$  的相关联程度的变量.一个直观的认识是,越是相关的任务,其权值矩阵  $w$  越相近.本文采用两种核函数来评估任务之间的相关联程度.包括高斯核函数(见公式(8))和多项式核函数(见公式(9));  $l(w_t, d_{t'})$  与公式(2)类似,表示 CRF 概率目标函数的对数形式,见公式(10).

$$\alpha_{t,t'} \triangleq \frac{1}{C} \exp\left(-\frac{\|w_t - w_{t'}\|^2}{2\sigma^2}\right) \quad (8)$$

$$\alpha_{t,t'} \triangleq \frac{1}{C} \frac{\langle w_t, w_{t'} \rangle^d}{\|w\|^d \cdot \|w\|^d} \quad (9)$$

$$l(w_t, d_{t'}) \triangleq \sum_{i \in N_n} \log P(y_{t',i} | x_{t',i}, w_t) \quad (10)$$

在公式(8)和公式(9)中,  $C$  是一个实数常数,用来控制任务之间相关联程度的量级;公式(8)中的  $\sigma$  是高斯核参数;公式(9)中的  $d$  是多项式核参数.

在公式(10)中,  $N_n$  表示任务  $t$  中的  $n$  个训练样例,为了表示简单,我们定义  $l_t(i, w) \triangleq \log P(y_{t,i} | x_{t,i}, w_t)$ . 最终,本文的多任务学习目标函数如公式(11)所示.

$$Obj(W, D) = \sum_{t,t' \in N_T} [\alpha_{t,t'} \sum_{i \in N_n} l_t(i, w_t)] - \sum_{t \in N_T} \frac{\|w_t\|^2}{2\sigma_t^2} \quad (11)$$

为了描述简单,我们引入一个  $T \times T$  的矩阵  $A$ .其中,  $A_{t,t'} \triangleq \alpha_{t,t'}$ . 本文采用一种启发式的算法求解任务之间的相关性矩阵  $A$ , 并采用交替式的方法优化模型的特征权值  $W$  (如下面的算法所示).首先初始化权值矩阵  $W$  (初始化为零矩阵)和相关性矩阵  $A$  (初始化为单位矩阵,表示任务只与自身存在相互关联关系).接着,通过公式(11)求解各个任务的模型特征权值向量  $w_t$ , 然后根据求解出的  $w_t$  运用公式(8)两两进行  $\alpha_{t,t'}$  的计算(即:更新相关程度矩阵  $A$ ).循环交替更新直至模型收敛并最终得到相关性矩阵  $A$  和特征权值矩阵  $W$ .

**算法.** 基于 CRF 的多任务学习算法.

输入:  $W \rightarrow 0, D, A$ ; 初始化特征权值矩阵  $W$  为零矩阵,初始化相关性矩阵  $A$  为单位矩阵;

for  $t \leftarrow 1$  to  $T$

    根据公式(11)分别计算特征权值矩阵  $w_t$

    for  $t \leftarrow 1$  to  $T$

        for  $t' \leftarrow 1$  to  $T$

            使用公式(8)或公式(9)更新  $A_{t,t'}$

            由  $A_{t,t'}$  和公式(11)更新权值矩阵  $w_t$

        反复迭代直至模型收敛



输出:收敛的相关性矩阵  $A$  和特征权值矩阵  $W$ .

## 4 实验结果及分析

### 4.1 数据与评价指标

本文采用 ACE 2005 中文语料进行实验,该语料共包含 633 个文档,内容涉及新闻、广播、博客、论坛和电话录音等.我们采用如下评测标准.即:

事件触发词识别/分类:一个事件触发词被正确识别当且仅当该事件触发词在文本的位置和类型与标准标注文档中的候选事件触发词的位置和类型完全匹配.

事件元素识别/分类:一个事件元素被正确识别当且仅当该事件元素在文本的位置和类型与标准标注文档中的候选事件元素的位置和类型完全匹配.

与大多数事件抽取研究的评价方法一致,我们采用准确率( $P$ )、召回率( $R$ )以及  $F_1$  值对实验结果进行评价.见公式(12)~公式(14).

$$P = \frac{\text{系统正确标注结果的总数}}{\text{系统标注结果的总数}} \quad (12)$$

$$R = \frac{\text{系统正确标注结果的总数}}{\text{语料中标准结果的总数}} \quad (13)$$

$$F_1 = \frac{2PR}{P+R} \quad (14)$$

### 4.2 实验设计

本文采用哈尔滨工业大学 LTP 平台(<http://www.ltp-cloud.com/>)进行实验预处理,包括分词、词性标注、命名实体识别和句法分析.与前人工作一致,本文也将 ACE 2005 中文语料提供的实体、值和时间信息当作已知信息使用.为使实验结果更具可比性,我们采用与基线方法<sup>[17-19]</sup>一致的语料划分方法.即:随机选取 567 篇文档作为训练集,66 篇文档作为测试集.此外,从训练集中随机选取 33 篇文档作为开发集.为了验证本文基于多任务学习的中文事件抽取联合模型(MTL-CRF)的有效性,将从如下 3 个方面进行论证.

- 1) 对基于 CRF 的事件抽取联合模型,分类训练能否改善混合训练的事件元素多标签冲突问题?
- 2) 联合模型能否提升分步模型的性能?
- 3) 引入多任务学习之后,联合模型的性能能否进一步提升?

为了验证问题 1),我们设计了两个系统:基于混合训练的事件抽取联合模型(MIX-CRF)(MIX-CRF、CLASS-CRF 以及 MTL-CRF 均为本文根据模型的特点所起的名字)和基于分类训练的事件抽取联合模型(CLASS-CRF),实验结果见第 4.3.2 节中的表 6;为了验证问题 2),我们将模型 CLASS-CRF 与目前性能最优的中文事件抽取分步模型(见表 4)进行对比,实验结果见第 4.3.3 节中的表 7.为了验证问题 3),本文将 CLASS-CRF 模型与基于多任务学习的中文事件抽取联合模型(MTL-CRF)进行对比,实验结果见第 4.3.3 节.另外,由于在条件随机场中,不同的特征模板会对实验结果产生较大影响,为了充分挖掘对事件抽取有用的特征,本文设计了 3 种不同的特征模板,分别为 Tempalte 1、Template 2、Tempalte 3,并在接下来的所有实验中选取性能最优的特征模板进行实验.3 种特征模板的实验结果见第 4.3.1 节中的表 5,所选取的具体特征如下.

#### • Template 1

词汇信息(6):当前词本身;当前词的词性;当前词的命名实体标签;前一词本身;前一词词性;前一词的命名实体标签.

句法信息(2):当前词在句法分析树中的位置;前一词在句法分析树中的位置.

最近实体信息(2):当前词的实体类型/子类型;前一词的实体类型/子类型.

#### • Tempalte 2

词汇信息(9):当前词本身;当前词的词性;当前词的命名实体标签;前一词本身;前一词词性;前一词的命名实

体标签;后一词本身;后一词词性;后一词的命名实体标签.

句法信息(3):当前词在句法分析树中的位置;前一词在句法分析树中的位置;后一词在句法分析树中的位置.

最近实体信息(3):当前词的实体类型/子类型;前一词的实体类型/子类型;后一词的实体类型/子类型.

#### • Tempalte 3

词汇信息(15):当前词本身;当前词的词性;当前词的命名实体标签;前两词本身;前两词词性;前两词的命名实体标签;后两词本身;后两词词性;后两词的命名实体标签.

句法信息(5):当前词在句法分析树中的位置;前两词在句法分析树中的位置;后两词在句法分析树中的位置.

最近实体信息(5):当前词的实体类型/子类型;前两词的实体类型/子类型;后两词的实体类型/子类型.

**Table 4** Comparison algorithms for Chinese event extraction

**表 4** 中文事件抽取对比算法

算法名称	算法原理简介
ECS+DC <sup>[17]</sup>	将组合语义和篇章一致性(discourse-consistency)相结合构建中文事件抽取模型,不仅能有效识别未登录的事件触发词,而且有助于模型对触发词进行类别分类
EMS+S <sup>[18]</sup>	在组合语义的基础上融合了形态结构和义原,有效地过滤组合语义方法形成的假触发词,平衡触发词正负样例的比例
ILP+MEMM+CRF <sup>[19]</sup>	采用 ILP(inter logic programming)推理架构将基于最大熵的事件触发词分类模型与基于 CRF 的事件触发词识别模型相结合,构建了触发词识别与触发词分类的中文事件抽取联合模型

为了验证核函数对任务相关度的影响,本文还分析了不同核函数(见公式(8)和公式(9))对 MTL-CRF 实验性能的影响,见第 4.4.1 节.并且给出了任务相关系数  $C$  值(公式(8)和公式(9))的讨论,见第 4.4.2 节.

### 4.3 实验结果

#### 4.3.1 不同 CRF 特征模板的性能比较

表 5 是本文 CLASS-CRF 事件抽取联合模型在选取不同特征模板时的实验结果.从表 5 中能够发现:当 CLASS-CRF 选取 Template 2 作为特征模板时实验性能最优.其事件触发词分类结果的  $F_1$  值达到 68%,比 Template 1 高 1.2%,比 Template 3 高 3.2%;事件元素分类结果的  $F_1$  值达到 51.8%,比 Template 1 高 2.8%,比 Template 3 高 0.7%.分析造成该情况的原因如下:(1) Template 2 比 Template 1 增加了后一词的词汇信息、句法信息以及最近实体信息,模型因为增加了适当的下文信息而使得性能提升;(2) 与 Template 2 相比,Template 3 虽然将前后词的依赖长度扩展为两个词,但造成了语料稀疏,反而有损模型性能.

**Table 5** Performance comparison of different CLASS-CRF feature templates

**表 5** CLASS-CRF 模型在不同特征模板下的实验性能对比

特征模板	性能	事件触发词分类			事件元素分类		
		$P(\%)$	$R(\%)$	$F_1$	$P(\%)$	$R(\%)$	$F_1$
Template 1		88.9	53.5	66.8	63.2	40.0	49.0
Template 2		87.2	55.9	68.0	65.1	43.0	51.8
Template 3		86.0	52.0	64.8	63.7	42.7	51.1

#### 4.3.2 事件元素的多标签情况对事件抽取性能的影响

为了解决事件元素的多标签问题,我们提出将基于 CRF 的事件抽取模型进行分类训练(CLASS-CRF). MIX-CRF 则将所有事件合在一起,训练一个统一的基于 CRF 的事件抽取联合模型.相比于 MIX-CRF,CLASS-CRF 事件触发词分类结果的  $F_1$  值提升了 14.1 个百分点,事件元素分类结果的  $F_1$  值提升了 6.8 个百分点.这说明在联合模型中,事件元素的多标签情况不容忽视.本文分析 CLASS-CRF 性能显著提升的原因有两点:(1) 一个事件提及中包含多个事件的情况约占总事件数目的 36.5%,分类训练将有助于这些事件标注样例的性能提升.(2) 由于在 MIX-CRF 中,事件元素多标签冲突的样例将被视为噪声,会给一个事件提及中只包含一个事件的情况造成影响.

**Table 6** Comparison of experimental performance of mixed CRF and classified CRF

**表 6** 混合 CRF 和分类 CRF 的实验性能对比

系统	性能	事件触发词分类			事件元素分类		
		P(%)	R(%)	F <sub>1</sub> (%)	P(%)	R(%)	F <sub>1</sub> (%)
MIX-CRF		68.3	44.6	53.9	57.5	37.0	45.0
CLASS-CRF		87.2	55.9	68.0	65.1	43.0	51.8

4.3.3 本文方法与基线方法的对比

通过观察表 7 可以发现:本文的 CLASS-CRF 模型的事件触发词分类结果的 F<sub>1</sub> 值优于 ECS+DC 模型(高 1.1%),劣于 EMS+S 模型(低 0.2%)和 ILP+MEMM+CRF 模型(低 2.2%);事件元素分类结果的 F<sub>1</sub> 值优于 ECS+DC 模型(高 1%),劣于 EMS+S 模型(低 0.2%)和 ILP+MEMM+CRF 模型(低 2.1%),达到了与基线方法可比较的水平.此外,本文的 MTL-CRF 模型的事件触发词分类结果的 F<sub>1</sub> 值优于 ECS+DC 模型(高 2%)、EMS+S 模型(高 0.7%)和 CLASS-CRF 模型(高 0.9%),低于 ILP+MEMM+CRF 模型(低 1.3%);事件元素分类结果的 F<sub>1</sub> 值优于 ECS+DC 模型(高 2.3%)、EMS+S 模型(高 1.1%)和 CLASS-CRF 模型(高 1.3%),略低于 ILP+MEMM+CRF 模型(低 0.8%).

**Table 7** Comparison of single-task model, multi-task model and baseline method

**表 7** 单任务模型和多任务模型与基线方法的实验对比

系统	性能	事件触发词分类			事件元素识别			事件元素分类		
		P(%)	R(%)	F <sub>1</sub>	P(%)	R(%)	F <sub>1</sub>	P(%)	R(%)	F <sub>1</sub>
ECS+DC		75.2	60.2	66.9	61.6	50.2	55.3	56.9	45.8	50.8
EMS+S		69.9	66.5	68.2	57.6	55.0	56.3	54.1	50.1	52.0
ILP+MEMM+CRF		71.4	68.9	70.2	59.1	57.2	58.1	55.8	52.1	53.9
CLASS CRF		87.2	55.9	68.0	69.8	46.0	55.5	65.1	43.0	51.8
MTL CRF		87.4	56.8	68.9	70.4	48.6	57.4	66.6	44.1	53.1

本文联合模型(CLASS-CRF 和 MTL-CRF)相比于基线方法拥有更高的准确率,但召回率却低于基线方法,接下来将从两方面对实验结果进行深入分析.

(1) 准确率方面:本文的 CLASS-CRF 模型及 MTL-CRF 模型在事件触发词分类和事件元素分类的准确率上都显著优于基线方法.这得益于如下几点:(a) 汉语是篇章驱动(discourse-driven)的语言<sup>[17]</sup>,不同的词在不同的上下文中往往具有不同的含义,很多假的事件触发词很难通过词性过滤或者形态结构过滤方法消除.(b) 事件抽取联合模型使事件触发词识别/分类和事件元素识别/分类同时进行,模型在进行事件触发词分类的过程中也能获取事件元素的分类信息,该信息将辅助完成事件触发词识别/分类.(c) 事件触发词分类的正收益也会通过联合模型传递到事件元素分类过程中,当模型的事件触发词分类的准确率提高了,也就代表着模型检测事件的准确率提高了,相应的事件元素分类的准确率也会随之提高.此外,MTL-CRF 模型相比于 CLASS-CRF,其准确率也有一定的提升,这也说明相关子类别事件确实可以起到相互增强的效果.

(2) 召回率方面:本文的 CLASS-CRF 模型及 MTL-CRF 模型在事件触发词分类和事件元素分类的召回率上都低于基线方法.主要有如下原因:(a) 由于联合模型构建了事件触发词识别/分类对于事件元素识别/分类的后向依赖关系,使得模型判定事件的“门槛”更高,很多正例事件由于没有事件元素或者部分事件元素缺失,使得模型很难将其判定为正例事件.同样,这种负收益也会通过联合模型构建的相互依赖关系传递到事件元素识别/分类中.(b) 另一方面,由于本文没有采用基线方法所使用的组合语义方法识别未登录的事件触发词,导致模型召回率不高.例如“中弹(injure)”“草菅人命(die)”“引火自焚(attack)”等词无法被本文模型识别为事件触发词.(c) 分词错误也是影响本文模型性能的很重要的因素,很多事件触发词和事件元素由于分词错误,导致模型无法准确识别其类别.例如:在测试语料中,“爆炸案”一词很难被准确地识别为 Attack 事件,原因就是分词工具将“爆炸”和“案”分在一起,而在标注文档中“爆炸”才是触发词.

(3) 相比于文献[19],本文在事件触发词分类和事件元素分类的 F<sub>1</sub> 值上都处于劣势.本文分析造成该情况的原因有如下两点:(a) 单字的事件触发词存在很强的歧义性,诸如“投”“中”“做”等.文献[19]采用基于依存句法分析的推理规则进行单字触发词的特殊处理.本文由于没有对单字触发词进行特殊处理,导致对单字触发词的

识别效果并不好,经统计,单字触发词的识别错误率高达 42.2%。(b) ACE 2005 中文语料中存在一些标注不合理的情况.文献[19]通过构建一些规则将其过滤,而在本文模型中,这些样例被视为模型的噪声,因而有损模型的性能.诸如下面两个样例:(A) 主要军港设立了 18 个环境检测站.(B) 广州等地设立了代表处.上述(A)、(B)两句中的“设立”表达涵义基本一致,但标注语料中仅将(A)句中“设立”标注为事件触发词,(B)句中未被标出.经本文统计,在测试集中共包含 11 个“设立”的正例事件,只有 1 个事件被本文模型准确识别为 Start-Org 事件.

#### 4.3.4 CLASS-CRF 模型与 MTL-CRF 模型在各事件大类下的性能对比

观察图 6 可以发现:MTL-CRF 模型相比于 CLASS-CRF 模型,总体性能有所提升.其事件触发词分类结果的  $F_1$  值在 Transaction 事件大类下提升约 20%,在 Business 事件大类下提升约 3%,在 Personnel 事件大类下提升约 4%.事件元素分类结果的  $F_1$  值在 Life 事件大类下提升约 6%,在 Transaction 事件大类下提升约 10%,在 Business 事件大类下提升约 1%,在 Personnel 事件大类下提升约 5%.对于性能没有提升的事件大类,本文分析造成该情况的原因主要有如下几点:(a) 标注的事件数目过于稀少,事件子类别之间几乎没有什么可以相互增强的信息(如:Justice 事件大类),(b) 事件元素的标签本身就不太相关联(如 Conflict 事件大类),(c) 选取的测试样本数太少,没有显现出 MTL-CRF 的性能优势(如 Contact 事件大类).

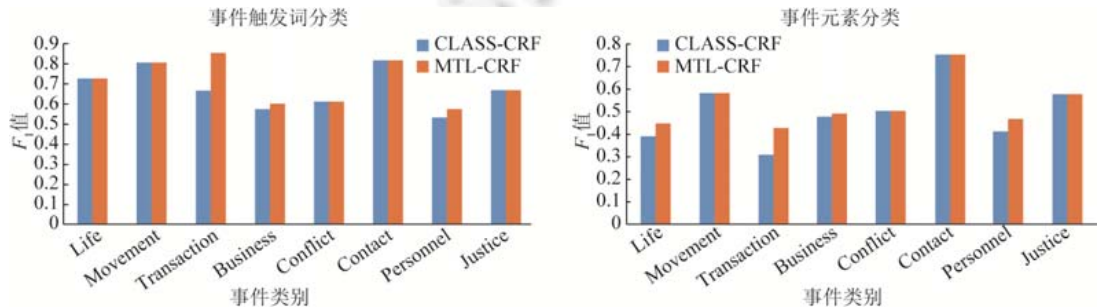


Fig.6 Comparison on performances of single-task and multi-task

图 6 单任务模型和多任务模型的性能对比

此外,对于性能提升的事件大类,本文分析原因如下:单任务事件抽取联合模型(CLASS-CRF)由于将语料进行分类训练,导致训练语料规模大幅缩小,使得最终模型包含的信息有限,极大地影响了模型的召回率.加入多任务学习模型后,由于相关任务共享模型所学信息,使得模型所包含的信息有所增长,能够有效地缓解语料稀疏问题并最终提升事件抽取模型的召回率.例如,测试样例“在警方驱散示威人群的过程中,至少有 5 个人受伤.”在 CLASS-CRF 模型中没有准确地将“警方”识别为 Agent 事件元素,而在 MTL-CRF 中则准确地将“警方”识别为 Agent 事件元素.本文通过观察语料发现,由于“警方”一词在 Injure 事件的训练语料中没有作为 Agent 事件元素出现的情况,因此,CLASS-CRF 模型无法在测试过程中将“警方”标注为 Agent 事件元素.而“警方”一词在 Die 事件的训练语料中作为 Agent 事件元素却出现了很多次,因此,MTL-CRF 模型通过信息共享机制,能够准确地将上述样例中的“警方”标注为 Agent 事件元素.

## 4.4 参数调整

### 4.4.1 关于核函数的讨论

在引入多任务学习增强相关子事件识别的过程中,本文分别采用高斯核函数以及多项式核函数来评估任务之间的相关程度(见公式(8)和公式(9)).图 7 所示为两种核函数分别在事件触发词识别和事件元素识别上的性能对比.整体上看,两种核函数对实验结果的影响并不明显,但是高斯核函数能够取得相对更优的实验结果.(a) 对于事件触发词分类,多项式核函数仅在 Justice 事件大类下的实验性能优于高斯核函数(高 1%左右);而在 Life 事件大类、Conflict 事件大类都劣于高斯核函数(分别低 2%左右和 3%左右).(b) 对于事件元素分类,高斯核函数仅在 Justice 类上实验性能略低于多项式核函数(低 0.2%左右),而在 Life 事件大类、Business 事件大类、Conflict 事件大类和 Personnel 事件大类都优于多项式核函数(分别高 3%左右、1.5%左右、1%左右、0.3%左右).

因此,本文采用高斯核函数作为事件子类别相关程度的度量.

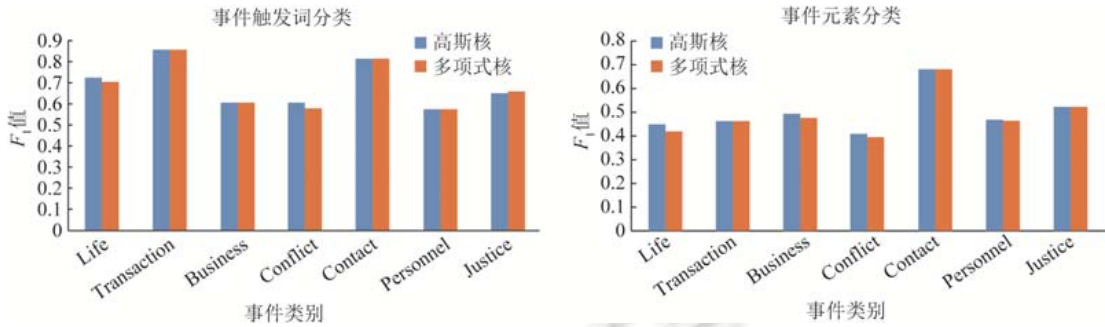


Fig.7 Discussion on kernel function

图7 核函数对系统性能的影响

#### 4.4.2 关于任务相关系数 $C$ 值的讨论

本小节将讨论在不同的事件大类下,当选择高斯核函数来度量任务之间的相关性时,任务相关系数  $C$  (公式(8)中的  $C$  值)的选择对实验结果的影响.观察图 8 可以发现:当  $C$  取值 1~20 时,MTL-CRF 模型的  $F_1$  值随着  $C$  值的增大而增大,当  $C$  值超过 20 时,性能反而出现一定程度的下滑.本文分析造成该情况有如下原因:(1) 当  $C$  取值为 1 时,公式(8)中的  $\alpha_{t,r}$  取值为  $[0,1]$ ,此时的实验效果并不好,说明该区间范围并不能很好地模拟同一事件大类下事件子类之间的相互关系.(2) 随着  $C$  值的增大,实验性能也随之提升,并且在  $C=20$  时性能达到最优,此时,  $\alpha_{t,r}$  取值范围为  $[0,0.05]$ ,说明该区间范围能够模拟事件子类之间的相互关系.(3) 当  $C$  值继续增大时,由于  $\alpha_{t,r}$  的取值范围进一步缩小,导致任务子类之间的相关性减弱,因此模型性能越来越趋向于单任务模型的性能.

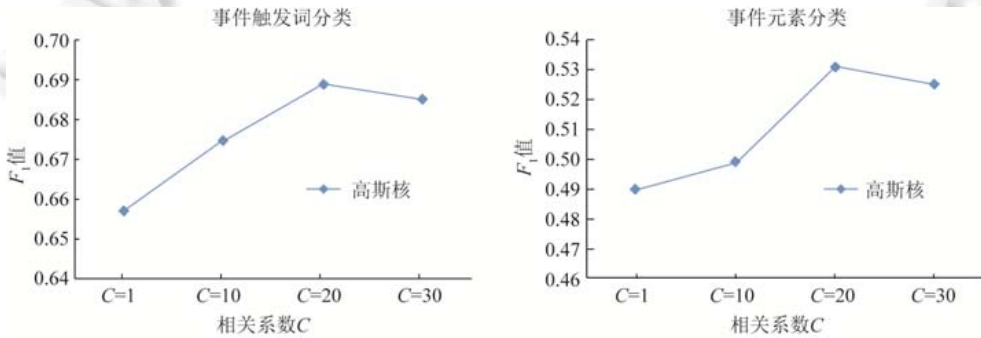


Fig.8 Discussion on correlation coefficient  $C$

图8 相关系数  $C$  对系统性能的影响

## 5 总结与展望

本文面向中文事件抽取提出采用基于 CRF 多任务学习的联合标注模型.通过总结前人方法存在的问题,从如下 3 个方面进行事件抽取模型的改进.(1) 为了解决管道事件抽取模型所带来的错误级联问题,以及事件触发词分类过程与事件元素分类过程缺乏交互的问题,本文将事件抽取看作是序列标注任务,对事件触发词和事件元素进行联合标注.(2) 为了解决联合标注模型带来的事件元素多标签问题,我们为每类事件分别训练一个基于 CRF 的事件抽取联合模型.(3) 为了解决分类训练后语料规模小和数据稀疏问题,提出采用多任务学习方法增强基于 CRF 的中文事件抽取联合模型,通过共享相关事件所学到的信息,有效缓解数据规模小和数据不平衡的问题.在 ACE 2005 中文语料上的实验结果表明了本文方法的有效性.

通过实验发现:在 ACE 2005 中文语料中存在很多事件类,其正样例数只有个位数的情况.诸如:Acquit 类事件、Pardon 类事件和 Execute 类事件等.对于这些事件,即使改进模型也很难取得令人满意的实验结果.因此,接

下来的工作中,本文将聚焦获取 ACE 2005 中文语料之外的信息或者事件抽取任务之外的信息.通过构建多任务学习模型,将其他与事件抽取相关的任务通过信息迁移的方式加入到 ACE 事件抽取任务中,以进一步改善事件抽取的性能.此外,尽管分类训练策略能够有效解决事件元素的多标签问题,但并不能完全杜绝该情况的发生,经过本文对 ACE 2005 中文语料进行统计,同一事件大类下的事件元素依然存在多标签的情况,占比约为 5.4%,同一事件子类下的事件元素也存在多标签的情况,占比约为 4.2%.如何完全消除事件元素的多标签情况也将成为本文接下来的研究重心.

**致谢** 衷心感谢审稿专家的悉心指导及本刊编辑的辛勤工作.

### References:

- [1] Li W, Wu M, Lu Q, Xu W, Yuan CF. Extractive summarization using inter-and intra-event relevance. In: Proc. of the ACL 2006. Stroudsburg: ACL, 2006. 369–376.
- [2] Frasinca F, Borsje J, Levering L. A semantic Web-based approach for building personalized news services. *Int'l Journal of E-Business Research (IJEBR)*, 2009,5(3):35–53. [doi: 10.4018/jebr.2009082103]
- [3] Borsje J, Hogenboom F, Frasinca F. Semi-automatic financial events discovery based on lexico-semantic patterns. *Int'l Journal of Web Engineering & Technology*, 2010,6(2):115–140. [doi: 10.1504/IJWET.2010.038242]
- [4] Ahn D. The stages of event extraction. In: Proc. of the Workshop on Annotating and Reasoning about Time and Events (ARTE 2006). Stroudsburg: ACL, 2006. 1–8.
- [5] Ji H, Grishman R. Refining event extraction through cross-document inference. In: Proc. of the ACL 2012. Stroudsburg: ACL, 2012. 835–844.
- [6] Liao SS, Grishman R. Using document level cross-event inference to improve event extraction. In: Proc. of the ACL 2010. Stroudsburg: ACL, 2010. 789–797.
- [7] Hong Y, Zhang JF, Ma B, Yao JM, Zhou GD, Zhu QM. Using cross-entity inference to improve event extraction. In: Proc. of the ACL 2011. Stroudsburg: ACL, 2011. 1127–1136.
- [8] Liao SS, Grishman R. Acquiring topic features to improve event extraction: In pre-selected and balanced collections. In: Proc. of the RANLP 2011. Stroudsburg: RANLP, 2011. 9–16.
- [9] Yang B, Mitchell T. Joint extraction of events and entities within a document context. In: Proc. of the NAACL 2016. Stroudsburg: NAACL, 2016. 289–299.
- [10] Li Q, Ji H, Huang L. Joint event extraction via structured prediction with global features. In: Proc. of the ACL 2013. Stroudsburg: ACL, 2013. 73–82.
- [11] Nguyen TH, Cho K, Grishman R. Joint event extraction via recurrent neural networks. In: Proc. of the NAACL 2016. Stroudsburg: NAACL, 2016. 300–309.
- [12] Chen YB, Xu LH, Liu K, Zeng DJ, Zhao J. Event extraction via dynamic multi-pooling convolutional neural networks. In: Proc. of the ACL 2015. Stroudsburg: ACL, 2015. 167–176.
- [13] Nguyen TH, Grishman R. Event detection and domain adaptation with convolutional neural networks. In: Proc. of the IJCNLP 2015. Stroudsburg: IJCNLP, 2015. 365–371.
- [14] Feng XC, Huang LF, Tang DY, Qin B, Ji H, Liu T. A language-independent neural network for event detection. In: Proc. of the ACL 2016. Stroudsburg: ACL, 2016. 66.
- [15] Liu SL, Chen YB, He SZ, Liu K, Zhao J. Leveraging FrameNet to improve automatic event detection. In: Proc. of the ACL 2016. Stroudsburg: ACL, 2016.
- [16] Chen Z, Ji H. Language specific issue and feature exploration in Chinese event extraction. In: Proc. of the HLT-NAACL 2009. Madison: Omnipress, 2009. 209–212.
- [17] Li PF, Zhou GD, Zhu QM, Hou LB. Employing compositional semantics and discourse consistency in Chinese event extraction. In: Proc. of the EMNLP 2012. Stroudsburg: ACL, 2012. 1006–1016.

- [18] Li PF, Zhou GD. Employing morphological structures and sememes for Chinese event extraction. In: Proc. of the COLING 2012. Mumbai: Indian Institute of Technology Bombay, 2012. 1619–1634.
- [19] Li PF, Zhu QM, Diao HJ, Zhou GD. Joint modeling of trigger identification and event type determination in Chinese event extraction. In: Proc. of the COLING 2012. Mumbai: Indian Institute of Technology Bombay, 2012. 1635–1652.
- [20] Li PF, Zhou GD, Zhu QM. Semantics-based joint model of Chinese event trigger extraction. Ruan Jian Xue Bao/Journal of Software, 2016,27(2):280–294 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4833.htm> [doi: 10.13328/j.cnki.jos.004833]
- [21] Zhao YY, Qin B, Che WX, Liu T. Research on Chinese event extraction. Journal of Chinese Information Processing, 2008,22(1): 3–8 (in Chinese with English abstract).
- [22] Fu JF, Liu ZT, Zhong ZM, Shan JF. Chinese event extraction based on feature weighting. Information Technology Journal, 2010,9:184–187. [doi: 10.3923/itj.2010.184.187]
- [23] Chen C, NG V. Joint modeling for Chinese event extraction with rich linguistic features. In: Proc. of the COLING 2012. Mumbai: Indian Institute of Technology Bombay, 2012. 529–544.
- [24] Miwa M, Thompson P, Korkontzelos I, Ananiadou S. Comparable study of event extraction in newswire and biomedical domains. In: Proc. of the COLING 2014. Dublin: Ireland, 2014. 2270–2279.
- [25] Mcclosky D, Surdeanu M, Manning CD. Event extraction as dependency parsing for BioNLP 2011. In: Proc. of the ACL 2011. Stroudsburg: ACL, 2011. 1626–1635.
- [26] Araki J, Mitamura T. Joint event trigger identification and event coreference resolution with structured perceptron. In: Proc. of the EMNLP 2015. Lisbon: Portugal, 2015. 2074–2080.
- [27] Lu W, Roth D. Automatic event extraction with structured preference modeling. In: Proc. of the ACL 2012. Stroudsburg: ACL, 2012. 835–844.
- [28] Ritter A, Mausam, Etzioni O, Clark S. Open domain event extraction from Twitter. In: Proc. of the ACM SIGKDD 2012. Stroudsburg: ACM SIGKDD, 2012. 1104–1112.
- [29] Li PF, Zhu QM, Zhou GD. Joint modeling of argument identification and role determination in Chinese event extraction with discourse-level information. In: Proc. of the IJCAI 2013. Stroudsburg: IJCAI, 2013. 2120–2126.
- [30] Caruana R. Multitask learning. Machine Learning, 1997,28(1):41–75. [doi: 10.1023/A:1007379606734]
- [31] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proc. of the ICML 2008. Stroudsburg: ICML, 2008. 160–167.
- [32] Liu X, Gao J, He X, *et al.* Representation learning using multi-task deep neural networks for semantic classification and information retrieval. In: Proc. of the HLT-NAACL 2015. Madison: Omnipress, 2015. 912–921.
- [33] Liu P, Qiu X, Huang X. Recurrent neural network for text classification with multi-task learning. arXiv Preprint arXiv: 1605.05101, 2016.
- [34] Liu P, Qiu X, Huang X. Deep multi-task learning with shared memory. arXiv Preprint arXiv: 1609.07222, 2016.
- [35] Daumé III H. Bayesian multitask learning with latent hierarchies. In: Proc. of the AUAI 2009. Stroudsburg: AUAI, 2009. 135–142.
- [36] Prettenhofer P, Stein B. Cross-language text classification using structural correspondence learning. In: Proc. of the ACL 2010. Stroudsburg: ACL, 2010. 1118–1127.
- [37] Sun X, Kashima H, Ueda N. Large-scale personalized human activity recognition using online multitask learning. IEEE Trans. on Knowledge and Data Engineering, 2013,25(11):2551–2563. [doi: 10.1109/TKDE.2012.246]
- [38] Lafferty J, McCallum A, Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proc. of the ICML 2001. Stroudsburg: ICML, 2001. 282–289.
- [39] Xu HY, Ren ZH, Shi J, Zhou H. An improved Chinese word segmentation method based on chain conditional random fields. Computer Applications and Software, 2016,33(12):211–213 (in Chinese with English abstract).
- [40] Qiu QQ, Miao DQ, Zhang ZF. Named entity recognition on Chinese Microblog. Computer Science, 2013,40(6):196–198 (in Chinese with English abstract). [doi: 10.3969/j.issn.1002-137X.2013.06.042]



## 附中文参考文献:

- [20] 李培峰,周国栋,朱巧明.基于语义的中文事件触发词抽取联合模型.软件学报,2016,27(2):280-294. <http://www.jos.org.cn/1000-9825/4833.htm> [doi: 10.13328/j.cnki.jos.004833]
- [21] 赵妍妍,秦兵,车万翔,刘挺.中文信息抽取技术研究.中文信息学报,2008,22(1):3-8.
- [39] 徐浩煜,任智慧,施俊,周晗.基于链式条件随机场的中文分词改进方法.计算机应用与软件,2016,33(12):211-213.
- [40] 邱泉清,苗夺谦,张志飞.中文微博命名实体识别.计算机科学,2013,40(6):196-198.



贺瑞芳(1979—),女,山西忻州人,博士,副教授,CCF 专业会员,主要研究领域为自然语言处理,社交媒体挖掘,机器学习.



段绍杨(1993—),男,硕士生,主要研究领域为中英文事件抽取,深度学习,条件随机场.