

基于语义的中文事件触发词抽取联合模型^{*}

李培峰, 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)

通讯作者: 李培峰, E-mail: pfli@suda.edu.cn



摘要: 中文事件触发词抽取是一项具有挑战性的任务. 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, 提出了基于语义的中文事件触发词抽取联合学习模型. 首先, 根据中文句子结构灵活和句法成分多省略的特点, 提出了基于模式匹配的核心论元和辅助论元抽取方法, 这两类论元可以较好地表示论元语义, 进一步提高中文事件触发词抽取性能; 其次, 根据同一文档中关联事件实例间存在的高度一致性, 构造了一个关联事件语义驱动的中文事件触发词识别和类型分配二维联合模型, 用于抽取贫信息事件实例. 在 ACE 2005 中文语料上的实验结果表明: 与现有最好的中文事件抽取系统相比, 所提出方法的性能得到了明显提升.

关键词: 中文触发词抽取; 论元语义; 关联事件语义; 联合学习模型; 核心论元

中图法分类号: TP391

中文引用格式: 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>

英文引用格式: 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). <http://www.jos.org.cn/1000-9825/4833.htm>

Semantics-Based Joint Model of Chinese Event Trigger Extraction

LI Pei-Feng, ZHOU Guo-Dong, ZHU Qiao-Ming

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

Abstract: Chinese event trigger extraction is a challenging task. To tackle the difficulties of obtaining the semantic information of event arguments and extracting those context-poor event mentions in Chinese event trigger extraction, this paper proposes a semantics-driven joint model to integrate the components of Chinese event trigger extraction. First, considering the nature of Chinese language (e.g., flexible sentence structure and ellipsis), it provides a pattern-based method to identify core arguments and supplement arguments to better represent argument semantics, and applies the method to improve the performance of Chinese trigger extraction. Secondly, regarding the consistency among relevant event mentions in a document or discourse, it introduces the semantics among relevant event mentions to formulate a 2-dimensional joint model of Chinese trigger detection and type allocation to extract those context-poor event mentions. Finally, it provides experimental results on the ACE 2005 Chinese corpus to show that the presented model significantly outperforms the state-of-the-art system.

Key words: Chinese trigger extraction; argument semantics; relevant event semantics; joint learning model; core arguments

作为信息抽取的一个子任务, 事件抽取(event extraction)主要从文本中抽取预先定义好的各种类型事件实例(event mention, 描述一个事件的句子)及其论元(argument, 事件的参与者和属性, 由实体实例组成, 是构成事件的基本要素). 事件抽取一般分为 2 个子任务: (事件)触发词(trigger, 用于标识事件的谓词, 一般动词和名词居多)

* 基金项目: 国家自然科学基金(61472265, 61331011); 软件新技术与产业化协同创新中心资助项目

Foundation item: National Natural Science Foundation of China (61472265, 61331011); Supporting Program of Collaborative Innovation Center of Novel Software Technology and Industrialization

收稿时间: 2014-09-19; 修改时间: 2015-02-09, 2015-03-16; 采用时间: 2015-03-24

抽取和论元抽取,如例 1 所示:

例 1:维和部队(A_1 :Artifact)25 日(A_2 :Time)乘飞机(A_3 :Vehicle)返回(EM_1 :Transport)安曼(A_4 :Destination).

触发词抽取主要从文本中抽取事件的触发词实例(根据 ACE 评测任务的定义,抽取事件触发词实例等同于抽取事件实例)并判定其事件类型(event type).从例 1 中可抽取一个由触发词“返回”触发、事件类型为 Transport(运输)的事件实例 EM_1 .论元抽取主要抽取特定类型事件的论元并判定其角色(role).从例 1 中可抽取事件实例 EM_1 的 4 个论元 $A_1\sim A_4$,并识别其角色分别为 Artifact(运送对象),Time(时间),Vehicle(交通工具)和 Destination(目的地).

由于事件抽取严重依赖于其前续环节(如实体识别、句法和依存分析等),且事件结构建立在语义而非语法基础上,导致了事件抽取系统的性能低下.目前,触发词抽取和论元抽取的 $F1$ 值一般分别在 50%~70%和 40%~60%左右^[1-3].其中,论元抽取性能低的主要原因是其前续环节触发词抽取性能低下.若假设所有的触发词实例都正确抽取,那么论元抽取的 $F1$ 值可达 85%以上.所以,事件抽取的关键是触发词抽取,这也是本文的研究重点.

当前,绝大多数的事件抽取研究集中在英文,由于中文本身特有的特点和难点,只有少数研究聚焦在中文事件抽取上.中文事件抽取,特别是中文触发词抽取,还存在两个主要问题:一是如何正确获取事件的论元语义;二是如何抽取那些缺乏足够有效识别信息的事件实例.

事件的核心是论元和触发词,它们的语义以及关系是事件抽取中最为重要的信息.由于论元语义信息难以获取,在英文中绝大多数句子级别的事件抽取系统还是沿用了语义角色标注(semantic role labeling,简称 SRL)的方法,充分利用句法信息来识别事件实例及其论元.中文是一种意合语言,其句子结构较为松散,句子成分搭配相对英文等拼音文字而言更加灵活,这些中文固有特点造成了句法特征在中文事件抽取中的效果不像在英文中那么明显.另外,事件的论元和角色不同于 SRL 中的论元和角色.SRL 论元表达了和谓词之间的语法关系,对应角色只有笼统的几种(如 $Arg_0\sim Arg_5, Arg_M$).而事件论元(下文提到的论元特指事件论元)则是事件的参与者或属性,对应的角色类别较多(上百种).而且,由于句子表述的多样性,论元和触发词之间的关系并不一定存在特定的语法结构.论元语义信息在触发词抽取中起了重要的作用,所以,如何针对中文行文特点从事件中获取更为有效、适合中文事件抽取的论元语义信息,是需要解决的一个关键问题.

另外,作为话题驱动语言,中文为了表述的连贯性和简洁性,部分句法成分缺省是一种常态.中文中存在的大量省略(特别是主语和宾语的省略)^[2,4],造成了部分事件论元信息缺失,使得那些缺乏足够有效识别信息(贫信息)的事件实例无法被抽取.所以,如何根据篇章内事件间的关系抽取贫信息中文事件实例,是需要解决的另一关键问题.

针对以上两个问题,本文提出了基于语义的中文事件触发词抽取联合学习模型.首先,根据中文句子结构灵活和句法成分多省略的特点,提出了基于模式匹配的核心论元和辅助论元抽取方法,使用这两类论元表示论元语义,帮助中文事件触发词抽取进一步提高性能;其次,根据同一文档中关联事件实例间存在的高度一致性,构造了一个关联事件语义驱动的中文事件触发词识别和类型分配二维联合模型,用于抽取贫信息事件实例.该模型根据事件的关联性,利用富信息的事件实例推导与之关联的贫信息的事件实例.在 ACE 2005 中文语料上的实验证明:与现有最好的中文事件抽取方法相比,本文方法的性能得到了明显的提升.

1 相关工作

近几年来,如何利用各种不同层次信息是事件抽取的主要研究方向.英文事件抽取研究开展相对较早,早期的研究主要偏重于模式匹配方法^[5-7],而近期的研究则倾向于机器学习方法.机器学习方法一般可分为句子级别事件抽取和文档/跨文档级别事件抽取两种.前者只考虑句子内部信息,包括词汇、短语和句法等不同层面^[8,9],后者是近期的研究重点,采用的方法包括引入了长距离特征的近似推理方法^[10]、基于聚类的事件和论元一致性推理方法^[11,12]、基于文档相关性的方法^[13]、结构预测方法^[3,14]和跨事件推理方法^[15]等.

目前,中文事件抽取还处于起步阶段,大多数研究主要从特征选择和触发词扩展两方面入手.

在特征选择方面,Chen 和 Ji^[16]根据中文触发词可能位于某个词内的特点,从词汇、句法、语义和相邻信息

等多个角度抽取事件的特征.Fu 等人^[17]提出了一个特征加权方法对中文事件的各种特征进行权重重新分配.Wang^[18]把各种特征,如项频度、句子位置和长度、标题词覆盖率、语义角色标注信息等组合起来,选择信息化最大的句子作为候选事件句.Chen 和 Ng^[19]把字符、语义角色标注、触发词概率、零指代、触发词一致性和论元一致性等特征用于中文事件抽取.赵妍妍等人^[20]对 Ahn^[8]的工作进行了特征方面的改进.付剑锋等人^[21]用依存分析来挖掘触发词与其他词之间的句法关系.本文前期工作^[22]提出利用主题模型对词语进行聚类,并利用类别信息来抽取中文事件.本文的前期工作^[23]还针对事件抽取论元角色缺失的现象,提出了基于跨事件的缺失事件角色填充方法.

在触发词扩展方面,本文的前期工作^[24]发现,未知事件触发词(不在训练语料中出现的触发词)是导致中文事件抽取系统性能低下的主要原因之一.目前,通过同义词或聚类方法进行触发词扩展是常用的方法.Chen 等人^[25]采用自举方法分别在英文和中文语料上进行事件抽取的联合训练,从而提高中文和英文的事件抽取性能.Ji^[26]从中英平行语料库入手,从英文语料中扩展新的中文触发词.Qin 等人^[27]则用同义词词林来扩展中文事件触发词.同义词扩展方法忽略了词汇的多义性,往往会降低准确率,从而影响整体性能.本文的前期工作^[24,28]根据中文词组的组合语言学,分别从动词构词结构和形态结构两个方面入手识别未知事件触发词,取得了良好的效果.

联合(学习)模型可以很好地协同两个或多个子任务:它不仅可把不确定信息从前面子任务传导到后面子任务,而且允许后面子任务中产生的有价值信息反馈到前面子任务.近年来,联合模型成为自然语言处理的一个研究热点,如联合中文分词和词性标注^[29,30]、联合词性标注和依存分析^[31]、联合句法分析和语义角色标注^[32]等.在事件抽取方面相关研究较少,而且多数应用于生物医学领域.Riedel 等人^[33]、Poon 等人^[34]利用马尔科夫逻辑构建联合学习框架来抽取生物医学中的事件实例.Riedel 等人^[35]提出了面向生物医学事件抽取的 3 个联合模型:第 1 个联合了触发词和论元的识别,第 2 个增加了在层次规则架构方面的一致性,第 3 个则把第 1 个和第 2 个集成用于捕获相同事件中不同论元的内在联系.另外,Do 等人^[36]构建了一个强调(事件-事件)对分类器和(事件-时间)对分类器一致性的联合模型.Li 等人^[3]提出了一个基于结构预测的事件抽取联合学习架构,并利用束搜索(beam search)来获得较优解.在中文事件抽取方面,本文的前期工作^[1]提出了一个结合了中文事件触发词识别和分类的联合模型,该模型结合了多种硬约束条件,提高了抽取系统的性能.本文的前期工作^[37]还面向事件论元抽取提出了利用论元在事件内和事件间一致性进行事件抽取的联合模型.

2 基准系统

触发词抽取由 2 个子任务组成,包括:1) 触发词识别,用于识别触发词实例(事件实例),如识别例 1 的触发词实例“返回”及其对应的事件实例 EM_1 ;2) 触发词类型分配,用于为识别的触发词实例分配事件类型,如分配 Transport(运输)为 EM_1 的事件类型.论元抽取同样包括 2 个子任务:1) 论元识别,用于识别某个特定类型事件实例的论元,如识别“维和部队官兵”、“25 日”、“飞机”和“安曼”为例 1 事件实例 EM_1 的论元;2) 论元角色分配,为每个论元分配角色,如分配角色 Destination(目的地)给 EM_1 的论元“安曼”等.

本文的前期工作^[1,24]在 Chen 等人^[16]的基础上优化了中文事件抽取的特征,被用作本文触发词抽取(包括触发词识别和触发词类型分配)的基准系统.其中,触发词识别和触发词类型分配均采用最大熵模型(maximum-entropy,简称 ME),包括以下特征集合:

- 词汇信息(6):触发词本身;触发词的词性;触发词的前一词+触发词;触发词+触发词的最后一词;前一词的词性+触发词词性;触发词的词性+后一词的词性.
- 句法信息(4):触发词在句法树中的深度;触发词所在短语的类型;触发词到其支配子句顶层的路径;触发词父节点的短语结构.
- 语义信息(2):触发词是否存在中文 PropBank 的谓词表中;触发词在中文同义词词林中编号.
- 最近实体信息(8):在句法树上离触发词最近的左/右实体及其类型;在物理距离上离触发词最近的左/右实体及其类型.

另外,采用本文前期工作^[2]实现的中文事件论元抽取系统作为论元抽取(包括论元识别和论元角色分配)的基准系统.该基准系统同样采用最大熵模型,并在此基础上利用全局优化模型进行论元推理.论元抽取不是本文的研究重点,仅使用基准系统来汇报结果,具体方法请参考文献[2].

3 论元语义的获取和表示

在基准系统的各个特征中,前 3 类用于捕获事件的触发词语义,最后一类“最近实体信息”则用于捕获事件的论元语义.由于论元抽取是触发词抽取的后续环节,基准系统只能采用一种近似方法来获取论元语义:在句法树上和物理距离上离触发词实例越近的实体,越可能是该触发词实例所触发事件的论元.很明显,由于中文句子在表达上相对自由以及在长句中后续子句一般缺省主语,利用该方法抽取的实体实际上很大一部分都不是该事件实例的论元,这造成了触发词识别和触发词类型分配性能的低下.但是,这些用于表示论元语义的特征还是很有用.在本文的实验中,“最近实体信息”特征可提高触发词识别的 $F1$ 值大约 9%.所以,如何在触发词抽取中充分挖掘论元语义,是提高中文事件触发词抽取性能的一个关键因素.

3.1 核心论元和辅助论元

一般情况下,论元对触发词的作用主要通过它在事件中扮演的角色来体现,那些和事件有着紧密关系的角色及其对应的论元在识别事件过程中的作用更加明显.一个完整事件可以表达成为 5W1H 形式(who(施事者谁),what(什么),whom(受事者谁),when(什么时候),where(什么地方),how(怎么样)),其中,What 的语义一般由触发词体现,其余 4W 分别对应事件中扮演不同角色的论元;1H 则用于辅助说明事件的一些细节信息(如宣判事件中的判决结果、攻击事件中的武器等).其中,When 和 Where 对应的角色绝大多数情况下为时间和地点,它们对应的论元一般不具有明显的事件类型区分度;只有 Who 和 Whom 对应的角色所填充的论元才具有鲜明的事件类别区分度.为此,本文把事件的参与者定义为事件的核心角色,核心角色回答了事件中核心的 2W(who,whom)问题,一般是事件的施事者和受事者.填充核心角色的论元称为核心论元.附录中的表 6 列出了 33 类事件对应的核心角色以及核心角色对应的论元类型(实体类型)集合.从表 6 可看出:绝大多数的核心角色和触发词构成了“Who do What to Whom”的形式,可以表达该事件的主要语义.

此外,1H 对应的角色也可鲜明地指出事件类型.例 2 中,两个句子具有相同触发词(免除)和核心论元(“公司”和“他”),但是句子 a)还有一个类型为 JOB(工作岗位)的实体“总经理”,该实体可为识别该句为 End- Position(去职)事件提供依据.

例 2:(a) 公司决定免除他担任的总经理职务;(b) 公司决定免除他欠公司的款项.

所以,本文另外定义了一个辅助角色,主要包括交通工具、职位、具体罪行、武器等.赋予辅助角色的论元称为辅助论元.附录中的表 7 给出了各类事件的辅助角色及其对应的论元类型集合.在 ACE 定义的 33 类事件中,23 类事件有辅助角色.

3.2 核心和辅助论元获取方法

论元抽取作为触发词抽取的后续环节,在事件触发词实例未抽取以及事件类型未知的情况下,无法抽取其核心/辅助论元.一个易想到的方法是 2 遍策略,其步骤如下:(1) 采用流水线模式依次运行事件抽取的 4 个子任务,自动获得触发词实例、事件类型、所有论元以及论元对应的角色;(2) 根据第 3.1 节的定义,从获取的论元中选择核心论元和辅助论元作为新特征加入触发词识别和触发词类型分配子任务中,再重新运行一遍这两个子任务,获取新的事件触发词实例.这种方法类似于联合学习方法.然而,本文的实验结果表明,该方法并不能提高中文事件抽取系统的性能.其原因有二:(1) 中文事件抽取中各个子任务的性能偏低,经过 4 层分类后产生大量级联错误;(2) 分类器采用的归纳学习(inductive learning)的方法同样也为那些假事件实例抽取了和真事件实例语义相近的论元,从而降低了分类性能.

本文从半监督或无监督的事件抽取方法中得到启发,采用模式匹配方法来推测事件实例的论元信息.事件抽取的模式匹配方法一般采用直推学习(transductive learning),根据定义好的事件模式抽取事件实例.该方法具

有准确率高、召回率低的特点.准确率高表示抽取的论元及其分配的角色绝大多数正确,正好适用于预测事件论元.另外,虽然其召回率较低,但大多数事件的核心论元不止一个,很多时候只要抽取其中部分核心论元即可.

本文用〈触发词实例,论元,触发词实例和论元间依存路径〉作为模式来表示一个事件中触发词实例和论元的关系,例如,例1中论元“维和部队官兵和触发词实例“返回”的模式为〈返回,维和部队官兵,nsbj→conj〉.针对每个候选触发词实例,它的核心和辅助论元的抽取方法如下:

- 1) 初步获取该候选触发词实例的事件类型.不同于英文触发词语义的多变,绝大多数中文触发词只对应一种特定的事件类型(95%以上).本文采用以下简单方法来识别触发词类型:a) 如果该候选触发词在训练语料中出现,则把训练语料中概率最大的类别赋给该候选触发词;b) 如果该候选触发词不在训练语料中出现,则根据 Li 等人^[28]的方法利用形态结构和核心词素(一个词中表达主要语义的词素,如“重伤”中的词素“伤”)来构建其相似触发词(和该触发词具有相同形态结构和核心词素的触发词)集合,再把该集合中概率最大的事件类型赋给该候选触发词实例.有关形态结构和核心词素的定义和获取方法,请参考文献^[28].
- 2) 通过实体类型匹配从事件句中获取该事件类型定义的核心和辅助角色对应的候选实体实例集合.
- 3) 针对每个候选实体实例,获取该实体实例和对应的候选触发词实例之间的依存路径,在测试集中构建测试集模式集合.
- 4) 从标注的训练集中抽取每个事件实例的核心和辅助论元,为每个抽取的论元构建模式(触发词实例,论元,触发词实例和论元间依存路径),得到训练集模式集合.
- 5) 针对每个测试集模式集合中的模式,计算其和训练集模式集合中的每个模式的相似度,并取最大值作为该模式的最终评分.本文采用与 Liao 等人^[38]相似的方法来计算模式相似度,如公式(1)所示.其中, tri_i 和 dp 分别表示模式 p 的三元组(触发词,论元,依存路径); $WSim$ 用于从 HowNet 获取两个触发词的语义相似度,计算方法参考文献^[39]; $ESim$ 判断 2 个论元的类型是否一致,如果一致则赋值 1,否则赋值一个小数字 0.1; $DSim$ 用于计算两条依存路径的相似度,本文采用夹角余弦法.另外, δ 是一个参数,本文同样按照 Liao 等人的方法^[38]设置为 0.6.

$$PSim(p_i, p_j) = \delta \times WSim(tri_i, tri_j) + (1 - \delta) \times ESim(a_i, a_j) \times DSIm(dp_i, dp_j) \quad (1)$$

- 6) 根据最终评分(高者优先)分别选择触发词左边核心论元、右边核心论元以及辅助论元.考虑到中文句子的主谓宾结构,所以选择触发词实例两边各一个核心论元作为候选.

3.3 论元语法的表示

在核心和辅助论元已抽取的情况下,还需考虑如何表示它们的语义,从而能够被机器学习方法所感知.对论元进行聚类,用类别来表示论元语义是一种常用方法.但是,这种方法很难控制类别数目,而且性能较低.为此,本文采用一种简单方法来表示论元的语义,即用论元的实体类别、子类别、实体本身等来表示不同层面的论元语义,这些特征可表示论元本身的语义和在事件中扮演的角色语义.为此,本文用以下特征来替换第 2 节中的特征类别“最近实体信息”:

- 论元语义(8):核心和辅助论元的中心词;核心和辅助论元的实体类型和子类型;中心词相对于触发词的位置(0-前;1-后);触发词的直接主语和直接宾语及其实体类型(如不是实体,用“NULL”表示其类别).

此外,本文根据触发词和论元的形态结构进一步为分类器抽取有效特征.根据本文前期工作在触发词扩展时的假设^[28].具有相同形态结构和相同核心词素的触发词触发相同类型的事件具有很高概率.本文把这个假设扩展到论元:具有相同形态结构、相同核心词素的论元在特定的事件中担任相同的角色同样也具有较高的概率.例如,核心词素是“弹”且形态结构为修饰结构(前面的词素修改后面的词素)的论元对应的角色分布比较集中,绝大多数是武器,如“子弹”、“炸弹”和“炮弹”等.在 ACE 2005 中文语料库上的统计显示,91%的该类论元填充的角色是“Instrument(工具)”.为此,本文提出了新的形态结构特征,用核心词素来聚合具有相似语义或相同类别的论元和触发词,具体特征如下所示.

- 核心词素(4):触发词和论元的核心词素及他们在 HowNet 中的义原(sememe).

为了把论元语义运用到中文触发词抽取中,本文把以上这些源自论元语义的信息作为分类特征加入到第2节的中文触发词抽取基准系统中,从而为该系统增加论元的语义信息,更加明确事件的论元语义。

4 中文触发词抽取二维联合模型

本节针对贫信息事件实例抽取困难问题,提出了根据同一文档中关联事件的语义一致性原理进行事件推理的方法,并利用联合模型来实现这种关联事件实例间的推理。

4.1 关联事件

在 ACE 2005 中文语料库上的统计表明:在新闻语料中,超过 55%的论元在句子层面缺失,需要通过篇章知识来填充^[2,37]。而大量论元的缺失,将造成很多贫信息事件实例无法抽取。例 3 中,攻击事件实例 EM1(惨案)因附近有足够有效识别信息(如攻击者和攻击工具)而易识别,而 EM2(惨案)因上下文缺乏有效识别信息而无法识别。

例 3:14 日 7 点多(A_1 :Time),闵应军(A_2 :Attacker)在陈棚村小学(A_3 :Place)用菜刀(A_4 :Instrument)制造了一起惨案(EM_1 :Attack)...在惨案(EM_2 :Attack)中,多名学生(A_5 :Victim)被砍伤(EM_3 :Injure)。

事件在其发生、发展过程中与外在的其他事件之间有着本源的语义关系,共同构成一个话题或文档。所以,话题/文档结构从语义角度可看作相关事件的层次化组织方式,表达了事件间的语义联系。话题内的事件存在高度关联性,使得它具有很强的内聚力,在语言表达上也存在很强的一致性。Grosz 等人^[40]和 Harris^[41]指出:处于同一上下文环境中的相同词汇具有相同或相似的含义,每个话题总是围绕某些实体(一般会有 1 个或多个主角)或核心事件来展开。因此,应用语义一致性理论,利用话题内同类事件的一致性以及关联事件的相关性,根据部分富信息事件实例来推导贫信息事件实例,可提高事件抽取系统性能。

为了有效利用事件之间的关联性进行事件推理,本文从触发词本身语义和事件间语义两个层面入手,定义了 3 种不同的关联事件:(1) 相同触发词事件;(2) 相似触发词事件;(3) 相关事件。

相同词汇在话题或篇章中往往具有相同语义。相同触发词事件是指文档中具有相同触发词词形的事件实例,如例 2 中, EM_1 和 EM_2 具有相同触发词词形“惨案”,属于相同触发词事件。由于中文中绝大多数触发词只触发一类事件,相同触发词事件也是相同事件类型事件,即,同指事件。

除了相同词汇具有相同或相似语义外,具有相同核心词素及其相同形态结构的词汇同样也具有相似语义。本文把该类事件称为相似触发词事件。例 4 中,2 个触发词实例“担任”和“出任”虽然词形各不相同,但它们具有相同核心词素“任”以及相同的形态结构。这两个触发词都对应相同的事件类型 Start-Position(任新职),具有语义上的一致性,属于相似触发词事件。很多时候,同一文档中的相似触发词事件同样也属于同指事件。

例 4:武绍祖 26 岁的时候就担任(EM_1)全国学联主席,后来又出任(EM_2)国务院副总理王震的秘书。

另外,根据话题结构理论,一个话题内的事件实例间存在相关性,这种相关性体现在发生时间前后和因果关系等方面。本文把这些经常同现的事件类型称为相关事件。在例 3 中,攻击事件实例 EM_1 和受伤事件实例 EM_3 经常一起出现在文档中,而且受伤事件往往由于攻击事件而产生,它们具有因果关系,属于相关事件。当一个文档中某个事件出现后,其相关事件出现的概率将大为提高。

4.2 基于关联事件的全局优化模型

根据关联事件进行事件推理,可根据已抽取的富信息事件实例来推导出贫信息事件实例,如,根据例 3 中的已知攻击事件 EM_1 推导出 EM_2 也是攻击事件。本文把它看作一个最优化问题来求解,联合模型是进行全局优化的常用方法。本文的前期工作^[1]针对中文事件触发词抽取构建了一个联合模型,把触发词类型分配的结果反馈到触发词识别,从而部分消除了级联错误带来的不利影响。本文将在此基础上通过加入基于关联事件的推理方法来构建一个二维的联合模型。

首先是构建触发词抽取两个子任务的联合,即,触发词识别和触发词类型分配联合模型。利用触发词识别器(trigger detector,简称 TD)和触发词类型分配器(trigger type allocator,简称 TTA)(即,本文的 2 个基准系统)分别计算每个候选触发词实例为真的置信度 P_{TD} 和作为特定事件类型的置信度 P_{TTA} 。其中,TD 是一个两类分类器,TTA

是一个多类分类器.联合模型的目的是计算候选触发词实例 ct_j 是 c_i 类型(所有事件类型($1 \leq i \leq 33$)和非事件($i=0$))的得分 $Score(c_i|ct_j)$:

$$Score(c_i | ct_j) = \begin{cases} (1-\gamma) \times f_{TD}(P_{TD}(c_{TD} = 1 | ct_j, T)) + \gamma \times f_{TTA}(P_{TTA}(c_i | ct_j, T)), & 1 \leq i \leq 33 \\ (1-\gamma) \times f_{TD}(P_{TD}(c_{TD} = 0 | ct_j, T)) + \gamma \times f_{TTA}(P_{TTA}(c_i | ct_j, T)), & i = 0 \end{cases} \quad (2)$$

其中, $f_{TD}(P_{TD})$ 和 $f_{TTA}(P_{TTA})$ 是根据置信度计算得分的函数; c_{TD} 是事件触发词识别的类别,1表示是事件触发词,0表示不是事件触发词; T 是训练语料; γ 是权重参数.联合模型的目标是求解最好的分类结果,即在篇章(D)中所有得分的最大化:

$$\arg \max_C \sum_{ct_j \in D} Score(c_i | ct_j) \quad (3)$$

其中, C 表示文档 D 中的所有候选触发词实例的分类结果.

如果把事件触发词识别和触发词类型分配看作一个横向联合的流程,那么这是一种一维的联合学习方法.其缺点是使用的分类器存在独立同分布假设(independent and identically distributed,简称 i.i.d),把所有的候选事件实例都孤立起来,不考虑同一话题/文档中事件之间的的关联性,无法推导出贫信息事件实例.

其次,本文在该一维联合模型中加入基于关联事件的联合学习机制,根据富信息事件实例来推导与之关联的贫信息事件实例,从而实现触发词抽取的全局最优.如果把关联事件实例间的联合学习当作一个纵向联合学习的过程,那么就构建了一个如图1所示的二维联合学习模型.一方面,不同的候选触发词实例 EM_i 在 TD 和 TTA 间进行联合学习,使得后续子任务 TTA 的结果可影响前面的子任务 TD(纵向虚线);另一方面,具有关联性的事件实例(如 EM_1, EM_2 和 EM_3)之间也可联合学习.假如 EM_1 是一个富信息事件实例, EM_2 和 EM_3 是贫信息事件实例.如果它们三者之间存在关联性并且 EM_1 是一个已知的事件实例,那么利用联合学习方法可根据 EM_1 推导出它的关联事件实例 EM_2 和 EM_3 (横向实线).该二维联合模型突破了以往仅在不同子任务中以任务为单位的联合学习方法,实现了同一文档中关联事件实例间的联合学习.而且,纵向和横向的联合学习可相互作用,把关联事件推理和机器学习方法有效结合,从而获得良好的识别效果.

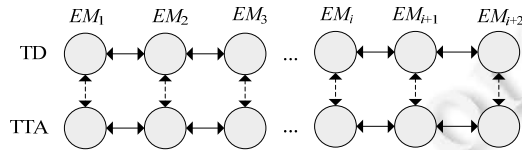


Fig.1 2-Dimension joint model of Chinese trigger detection and type allocation

图1 中文触发词识别和触发词类型分配二维联合模型

4.3 中文触发词识别和类型分配联合模型

为了更好地描述模型,首先定义表1所示变量.

Table 1 Pre-Defined variables for the joint model

表1 为联合模型预定义的变量

变量	说明
$t_{(i,j)}$	第 i 个触发词 W_i 的第 j 个实例.一个触发词在一个文档中可能存在多个实例
R	事件类型集合, r_j 是其中的一个事件类型. ACE 定义了 33 类不同事件
$p_{TD}(1 t_{(i,j)})$	$t_{(i,j)}$ 是候选触发词实例为真的置信度,其值来自 TD
$p_{TTA}(r_k t_{(i,j)})$	$t_{(i,j)}$ 是 r_k 类型事件触发词实例为真的置信度,其值来自 TTA
$c_{(i,j)}^{TD} / c_{(i,j)}^{-TD}$	$t_{(i,j)}$ 是/不是事件触发词实例的成本
$c_{(i,j,k)}^{TTA} / c_{(i,j,k)}^{-TTA}$	$t_{(i,j)}$ 是/不是 r_k 类型事件触发词实例的成本
$x_{(i,j)}$	指示触发词 $t_{(i,j)}$ 是不是事件触发词实例(0-不是;1-是)
$y_{(i,j,k)}$	指示触发词 $t_{(i,j)}$ 是不是 r_k 类型事件触发词实例(0-不是;1-是)
γ	TD 和 TTA 的权重比例,取值在 0-1 之间.

另外,类似本文前期工作^[1],定义了如下的成本函数:

$$c_{(i,j)}^{TD} = \log(p_{TD}(1 | t_{(i,j)})) \quad (4)$$

$$\bar{c}_{(i,j)}^{TD} = \log(1 - p_{TD}(1 | t_{(i,j)})) \quad (5)$$

$$c_{(i,j,k)}^{TTA} = \log(p_{TTA}(r_k | t_{(i,j)})) \quad (6)$$

$$\bar{c}_{(i,j,k)}^{TTA} = \log(1 - p_{TTA}(r_k | t_{(i,j)})) \quad (7)$$

本文采用整形线性规划(integer linear programming,简称 ILP)来解全局最优化问题.ILP 是一种基于约束推理的数学方法,它通过最小化或最大化满足一系列约束条件的目标函数,从而找到最优解.ILP 已被应用到很多自然语言处理中^[1,36].本文采用的 ILP 推理框架的目标函数如公式(8)所示,其中, D 表示文档中的触发词集合:

$$\max \sum_{W_i \in D} \sum_{t_{(i,j)} \in W_i} \left((1-\gamma) \times (c_{(i,j)}^{TD} \times x_{(i,j)} + \bar{c}_{(i,j)}^{TD} \times (1-x_{(i,j)})) + \gamma \times \sum_{W_i \in D} \sum_{t_{(i,j)} \in W_i} \sum_{1 \leq k \leq 33} (c_{(i,j,k)}^{TTA} \times y_{(i,j,k)} + \bar{c}_{(i,j,k)}^{TTA} \times (1-y_{(i,j,k)})) \right) \quad (8)$$

s.t.

$$x_{(i,j)} \in \{0,1\}, \forall W_i \in D \wedge t_{(i,j)} \in W_i \quad (9)$$

$$y_{(i,j,k)} \in \{0,1\}, \forall W_i \in D \wedge t_{(i,j)} \in W_i \wedge 1 \leq k \leq 33 \quad (10)$$

为了保证触发词识别器 TD 和触发词类型分配器 TTA 的一致性,本文提出了以下约束条件:

- 1) 如果一个触发词实例被 TTA 识别为 r_k 类型触发词,那么该实例必须被 TD 识别为事件触发词:

$$x_{(i,j)} \geq y_{(i,j,k)}, \forall W_i \in D \wedge t_{(i,j)} \in W_i \wedge 1 \leq k \leq 33 \quad (11)$$

- 2) 如果一个候选触发词实例被 TD 识别为真,那么该实例必须被 TTA 识别为某事件类型的触发词:

$$x_{(i,j)} = \sum_{1 \leq k \leq 33} y_{(i,j,k)}, \forall W_i \in D \wedge t_{(i,j)} \in W_i \quad (12)$$

4.4 基于关联事件的联合学习

根据第 3.1 节定义的 3 种关联事件,本文提出不同的约束条件进行事件触发词推理.本文的前期工作^[1]强制具有相同触发词的实例在文档内保持文档一致性(概率可达 90%左右),这导致了较多的错误发生.在 ACE 2005 中文语料上的统计表明:在文档中两个具有邻接关系(在文中相邻出现)的相同触发词实例往往在事件类别上具有更高的一致性,高达 95%以上.因此,本文提出基于邻接一致性的推理方法,通过强制具有相邻关系的相同词形触发词实例在事件类型上保持一致,从而根据富信息事件实例来推导贫信息事件实例.

如果强制所有的触发词实例都具有邻接一致性,约束条件具有的传递性会导致邻接一致性最终变成文档一致性.为此,本文利用条件约束来实现邻接一致性.即,约束条件在某种条件下才成立.从而打破约束条件的传递性,实现邻接一致性约束.相邻的相同触发词实例保持一致性的条件可归纳为 3 个方面:1) 触发词;2) 位置信息;3) 论元.首先,不同触发词具有不同概率的邻接一致性,图 2 列出了 ACE 2005 中文语料库中最高频的 6 个触发词的邻接一致性概率;其次,位置信息也可作为是否要保持一致性的标志,一般情况下,在一个句子(特别是长句)中先后出现的同一触发词的不同实例具有高度一致性;最后,事件句中具有相同、相似或同指论元也是保持一致性的标志.所以,本文提出了以下的约束条件:

$$y_{(i,j,k)} = y_{(i,j+1,k)}, \forall W_i \in D \wedge t_{(i,j)}, t_{(i,j+1)} \in W_i \wedge 1 \leq k \leq 33 \wedge (t_{(i,j)}, t_{(i,j+1)} \in S_m \vee P_{co}(W_i) > \alpha \vee \text{SimArg}(t_{(i,j)}, t_{(i,j+1)})) \quad (13)$$

其中, S_m 表示文档第 m 个句子中的所有触发词实例集合; $P_{co}(W_i)$ 表示触发词 W_i 的邻接一致性先验概率; α 是一个阈值,使得邻接一致性概率高的触发词才需要保持一致性; $\text{SimArg}(t_{(i,j)}, t_{(i,j+1)})$ 计算相邻触发词 $t_{(i,j)}$ 和 $t_{(i,j+1)}$ 是否具有相同和相似的实体,再利用语料中标注的实体指代链判断是否具有相同论元.只要 3 个条件中的某一个成立,就增加约束条件(13),否则不增加.

同样,在文档中两个邻接的相似触发词实例在事件类别上也具有很高的一致性,同样可用于事件推理.首先,利用本文前期工作^[28]的方法,用 $HM(tt_{(i,j)})$ 和 $Morph(tt_{(i,j)})$ 分别识别文档中的每个触发词的核心词素和形态结构;然后,把具有相同核心词素和形态结构的触发词实例根据在文中出现的先后次序构建相似触发词链集合 CC.其中,第 i 条链 cc_i 上的第 j 个触发词实例用 $tt_{(i,j)}$ 表示,用 $z_{(i,j,k)}$ 来表示 $tt_{(i,j)}$ 是否是一个 r_k 类型的事件实例(0-不是,1-是).利用触发词实例编号的对应关系,很容易把指示器 z 映射到公式(8)中的指示器 y .采用与相同触发词

类似的方法来保证其一致性:

$$z_{(i,j,k)} = z_{(i,j+1,k)}, \forall cc_i \in CC \wedge tt_{(i,j)} \in cc_i \wedge 1 \leq k \leq 33 \wedge HM(tt_{(i,j)}) = HM(tt_{(i,j+1)}) \wedge Morph(tt_{(i,j)}) = Morph(tt_{(i,j+1)}) \wedge (tt_{(i,j)}, tt_{(i,j+1)}) \in S_m \vee P_{co}(HM(tt_{(i,j)})) > \beta \vee SimArg(tt_{(i,j)}, tt_{(i,j+1)}) \quad (14)$$

其中, $P_{co}(HM(tt_{(i,j)}))$ 表示触发词实例 $tt_{(i,j)}$ 的核心词素的邻接一致性先验概率, β 是一个阈值.

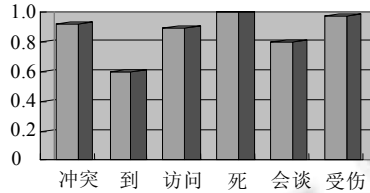


Fig.2 Probabilities of adjacent consistency of top 6 frequent triggers

图 2 出现频率最高的 6 个触发词的邻接一致性概率

不同类型事件出现在同一个文档时,往往也具有较强的一致性.例如,在 ACE 2005 中文语料上中攻击事件和死亡事件、死亡事件和受伤事件同现概率超过 70%,在文档中的一致性高达 93%以上.所以,利用相关事件一致性同样可根据富信息事件实例推导贫信息事件实例.

首先,根据训练集中的已知触发词集合从训练集文档中抽取所有事件实例,统计不同类型事件相互之间的一致性概率,得到 $33 \times 32 / 2 = 528$ 个不同的事件对一致性概率;然后,把这些一致性概率分成 3 个集合:高概率集合、中概率集合和低概率集合.具体的约束方法如下:对于高概率集合中的事件类型对,强制其在文档内的一致性;对于中概率集合中的事件类型对,强制其在长句内(内有逗号)或两个邻接短句(无逗号句)内的一致性;对于低概率集合中的事件类型对,不做任何约束处理.具体约束条件如下:

$$Y_{(i1,j1,k1)} = Y_{(i2,j2,k2)}, \forall W_{i1}, W_{i2} \in D \wedge t_{(i1,j1)} \in W_{i1} \wedge t_{(i2,j2)} \in W_{i2} \wedge 1 \leq k1, k2 \leq 33 \wedge (P_{cp}(k1, k2) \geq \theta \vee \theta > P_{cp}(k1, k2) \geq \lambda \wedge t_{(i1,j1)}, t_{(i2,j2)} \in LS) \quad (15)$$

其中, $P_{cp}(k1, k2)$ 为事件类型 $k1$ 和 $k2$ 的事件对的一致性概率; LS 为长句或两个邻接短句内所有触发词实例集合; θ 和 λ 是参数,用于限制一致性概率的范围.

5 实验和讨论

本节首先介绍了实验环境及其评测标准,然后给出了利用论元语义及其关联事件语义进行事件抽取的实验结果,最后对实验结果进行了深入分析.

5.1 实验环境

本文实验基于 ACE 2005 中文语料库,该语料库由新闻、广播、对话、博客、论坛和电话录音等组成,包括 633 个文档,标注了 8 个大类和 33 个子类的事件实例.和以往的研究一样,本文只考虑 33 个事件子类型,把识别子类型作为本文的任务(本文提及的事件类型特指其子类型).由于测试语料的不同选择会导致实验结果存在较大波动,为了获取更加正确的性能评估,不同于以往大多数研究选择 10% 的语料作为测试集、剩余的语料作为训练和开发集,本文采用 10-折交叉验证方法,把实验数据分成 10 份,每次用 1 份作为测试集,其余作为训练集.

和通常的信息抽取任务一样,本文通过准确率 $P(\text{precision})$ 、召回率 $R(\text{recall})$ 和 $F1$ 值来汇报实验性能.另外,和以往研究一样,也使用 ACE 2005 中文语料库标注的实体、值和时间等标注信息以及采用 Chen 等人^[16]、Li 等人^[1]的评测方法.

- 触发词识别:如果一个触发词实例的位置和标注文档一致,则该触发词实例识别正确.
- 触发词类型分配:如果一个触发词实例的位置和事件类型和标注文档一致,则该触发词实例的类型分配正确.
- 论元识别:如果一个论元对应事件类型和位置和标注文档一致,则该论元识别正确.

- 论元角色分配:如果一个论元对应事件类型、位置以及角色和标注文档一致,则该角色分配正确。

另外,本文采用 ICTCLAS 作为中文分词工具,Berkeley Parser 和 Stanford Parser 用于获取句法和依存信息;Stanford Classifier 和 Ip_solve 用于进行分类和全局优化。

5.2 实验结果

表 2 给出了基准系统、加入了论元语义以及二维联合模型在事件抽取 4 个子任务上的正确率 P 、召回率 R 和 $F1$ 值。其中,第 4.3 节和第 4.4 节中的参数 $\gamma, \alpha, \beta, \theta$ 和 λ 根据其中的 1 份测试集调优得到,分别设置为 0.4, 0.8, 0.9, 0.9 和 0.6。

Table 2 Comparison on performances of four subtasks of Chinese event extraction

表 2 中文事件抽取 4 个子任务的总体性能对比

系统/人工	触发词识别			触发词类型分配			论元识别			论元角色分配		
	P (%)	R (%)	$F1$	P (%)	R (%)	$F1$	P (%)	R (%)	$F1$	P (%)	R (%)	$F1$
基准系统	70.1	61.6	65.6	66.8	59.0	62.6	56.7	50.2	53.3	52.6	47.2	49.7
基准系统+论元语义	82.4	58.9	68.7	78.3	56.5	65.5	66.4	47.9	55.7	61.8	44.6	51.8
二维联合模型	78.7	69.0	73.5	74.8	65.8	70.0	63.5	56.0	59.5	58.8	52.1	55.2
标注者 1(不熟悉)	63.3	62.9	63.1	61.7	59.5	60.6	64.6	54.1	58.9	60.9	48.2	53.8
标注者 2(熟悉)	72.6	74.3	73.4	69.1	70.2	69.6	71.5	65.9	68.6	66.4	54.6	59.9
IAA	45.8	42.9	44.3	45.3	42.5	43.8	60.4	49.7	54.5	55.1	45.9	50.1

和基准系统相比,加入了论元语义可分别提高触发词识别、触发词类型分配、论元识别和论元角色分配的 $F1$ 值 3.1%, 2.9%, 2.4% 和 2.1%。这主要得益于它们在准确率上的大幅提高(依次为 12.3%, 11.5%, 9.7% 和 9.2%)。同样,二维联合模型在这 4 个子任务上的 $F1$ 值分别再度提高了 4.8%, 4.5%, 3.8% 和 3.4%, 这主要得益于在召回率上的大幅提高(依次为 10.1%, 9.3%, 8.1% 和 7.5%)。实验结果表明:本文提出的方法可较为显著地提高中文事件抽取各个子任务的性能,验证了其有效性。

论元语义带来的高准确率和联合模型(主要是关联事件语义的引入)带来的高召回率,说明它们在中文触发词抽取中具有很强的互补性。分析实验数据可发现:那些没有成功抽取核心论元的事件实例和那些贫信息事件实例,大多数没有被分类器所识别(详细分析参见第 5.3 节)。而基于关联事件的二维联合模型正好可弥补这个缺陷,它根据已经抽取的富信息事件实例推导与之关联的贫信息事件实例,从而提高了整个系统的召回率。

基准系统的最大熵模型采用通用迭代算法 GIS(generalized iterative scaling)求解,其时间复杂度为 $O(s \times p \times |F|)$ 。其中, s 是训练集合的大小, p 是预测类别的数量, $|F|$ 是特征集合的数量。而本文的二维联合模型则采用 ILP 求解,这是一个 NP 问题。但本文采用 2 值变量(0 或 1),所以本质上是一个 0-1 线性规划问题。本模型中所有的约束都是线性条件,寻找最优解的平均时间复杂度为 $O(2^n \times n \times m)$ 。其中, n 是变量的个数, m 是约束条件的个数。为了降低二维联合模型的解空间,本文一方面利用 ILP 工具提供的分支界定(branch and bound)算法进行解码;另一方面以文档为单位进行求解,通过降低每次求解的变量数目 n 来降低时间复杂度。另外,由于每个文档的事件实例个数平均在 6 个左右,所以 n 值一般不会很大。本文的实验运行在一个具有 16GB 内存 Intel i7 CPU 的计算机上,其中,基准系统耗时 67.5s,基准系统+论元语义耗时 71.3s,二维联合模型耗时 98.2s。其中,每个文档的最优化求解平均耗时均小于 0.1s。

10-折交叉验证实验结果表明,中文事件抽取在不同数据上的结果存在较大差异,如在触发词识别中,最好和最差的 $F1$ 值间的差距达到 11.9%(基准系统:最高 73.0% 和最低 61.9%)。这说明利用交叉验证方法能够避免语料选择的影响,使得实验结果更加客观。分析不同测试集的结果,在分类特征中加入论元语义信息后,在 9 个测试集上性能都有显著提高,仅在 1 个测试集上, $F1$ 值的提高小于 1%。分析这个测试集,主要的问题是语料的不平衡性。由于语料随机划分成 10 份,导致某些事件类型的实例大量存在于测试集中,而在训练集中明显不足。本文提出的二维联合模型在所有的测试集上均表现良好,其中,触发词识别的 $F1$ 值提高在 3%~6% 之间。

另外,2 个研究生重新标注了其中一份测试集。其中,标注者 1 不熟悉事件抽取,在花费半天时间阅读了标注说明文档后开始标注工作;标注者 2 熟悉事件抽取,在该领域有近 2 年的研究工作。表 2 的结果表明,人工标注的

$F1$ 值也仅仅在 50%~70%之间.这从另一方面说明中文事件抽取是一个具有挑战性的任务.标注者间的一致性 IAA(inter-annotator agreement)在触发词识别和类型分配上小于 45%.这不是标注质量问题,因为 Ji 和 Grishman^[14]在英文上人工标注结果的 IAA 也只有 40%左右.这种不一致性和较差的人工标注结果主要来自 2 个因素:(1) 中文中长句比较普遍,一个长句中往往存在多个事件实例,导致标注者容易遗漏部分实例;(2) 标注说明文档不够清晰,存在较多的歧义.另外,人工标注结果同样说明,在触发词识别和类型分配之间、论元识别和角色分配之间的性能差距较小.在触发词正确识别和论元正确识别的基础上,触发词类型分配和论元角色分配的准确率分别可高达 95%和 93%以上.这种现象也表明,触发词识别和论元识别是中文事件抽取中的 2 个最为重要环节.

本文同样在 ACE 2005 英文语料上验证了本方法的有效性,其中,基准系统的特征集合和实验环境与文献 [3]一致.由于论元语义抽取方法和相似触发词事件联合学习需要用到中文词语的核心词素,这在英文中无法实现.所以,本文仅把二维联合学习模型(不包括相似触发词事件联合学习)应用到英文中,实验结果见表 3.与英文基准系统相比,二维联合模型在 4 个子任务中分别大幅提高了 $F1$ 值,主要得益于在召回率上的大幅提高.与当前最好的英文事件抽取系统^[3]相比,在缺少相似触发词事件联合学习的情况下,触发词抽取的性能略优,而论元识别的性能略差.其中,论元识别性能比 Li 等人^[3]的系统差的原因是:该系统实际上主要专注于论元抽取,在这个环节提出了很多适合英文的全局优化方法.

Table 3 Comparison on performances of four subtasks of English event extraction

表 3 英文事件抽取 4 个子任务的总体性能对比

系统	触发词识别			触发词类型分配			论元识别			论元角色分配		
	P (%)	R (%)	$F1$	P (%)	R (%)	$F1$	P (%)	R (%)	$F1$	P (%)	R (%)	$F1$
基准系统	76.2	60.5	67.4	74.5	59.1	65.9	74.1	37.4	49.7	65.4	33.1	43.9
二维联合模型	73.5	68.7	71.0	70.9	65.4	68.0	64.7	50.3	56.6	60.1	46.2	52.2
Li ^[3] 的系统	76.9	65.0	70.4	73.7	62.3	67.5	69.8	47.9	56.8	64.7	44.4	52.7

5.3 论元语义对中文事件触发词抽取的作用

表 2 的结果表明,本文提出的论元语义抽取方法具有准确率高、召回率相对偏低的特点.其中,事件触发词抽取准确率的提高主要受益于高准确率的核心/辅助论元抽取方法.在本文的实验中,第 3.2 节中的核心和辅助论元获取方法的 $F1$ 值为 76.3%(P :88.4%, R :67.1%,不考虑那些没有抽取的事件实例和抽取的假事件实例).其中,58.9%事件实例的核心论元完全抽取正确,81.4%事件实例部分论元抽取正确(至少成功抽取一个以上).另外,论元语义信息的加入会影响中文触发词抽取的召回率,主要原因是大约 15%的事件实例没有抽取任何核心论元.这些没有抽取核心论元的事件实例以及部分贫信息事件实例基本上没有新加入的论元语义特征,所以大多数没有被分类器所识别,这是导致低召回率的主要原因.为了进一步验证论元语义的作用,本文做了如下实验:把测试集中所有事件实例的论元用标注文档中正确的核心论元替代(即,表示核心论元抽取完全正确).实验结果表明,事件触发词识别的 $F1$ 值可提高 11%左右.这同样验证了论元语义在事件触发词抽取中的巨大作用.

为了验证论元语义的不同表示方法对中文触发词识别性能的影响,见表 4,本文针对第 3.3 节提出的 6 个论元语义特征(集合)分别验证了其在触发词识别方面对性能提高的贡献度.从表 4 可以看出,每个特征均提高了 $F1$ 值.其中,论元中心词、论元类型、主语和宾语对中文事件触发词识别的影响最大,分别提高了 $F1$ 值达到 1.4%, 0.5%和 0.5%.这也符合一般常识,因为它们是表示论元语义的最主要方式.

对实验结果做进一步分析可知:除了分类器本身和核心/辅助论元抽取方法的性能影响之外,加入了论元语义的基准系统的错误主要来自以下 2 个方面:首先,测试集的随机生成造成某些类别事件实例在训练集中的正负比例严重失衡,降低了召回率,例如在本实验的第 2 层交叉验证中,触发词“设立”和“担任”在训练集合中的真假比例分别是 1:10 和 9:26,而在测试集中则为 7:1 和 5:1,这两个触发词的比例失衡就导致了大约 3%的召回率损失;其次,ACE 2005 中文语料库的标注不一致也导致了错误发生,如“战争”是攻击事件的一个常用触发词,在语料中共有 11 处出现了“朝鲜战争”和“海湾战争”,但是只有 4 处标注了攻击事件,其他的 7 处则没有标注.在仔

细分析了那些识别错误的事件实例后,大约有 12%左右的假事件实例实际上可以标注为真.表 2 中较低的 IAA 同样也说明了这个问题.

Table 4 Contributions of the different argument semantic features in Chinese trigger detection (incremental features)

表 4 论元语义特征在中文触发词识别中的贡献(特征递增)

特征	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i>
基准系统	70.1	61.6	65.6
+论元中心词	+9.3	-3.7	+1.4
+论元类型(包括子类型)	+2.1	-0.4	+0.5
+论元位置	+0.1	+0.1	+0.1
+主语和宾语	+0.3	+0.6	+0.5
+论元核心词素及其义原	+0.3	+0.3	+0.3
+触发词核心词素及其义原	+0.2	+0.4	+0.3

5.4 关联事件语义对中文事件触发词抽取的作用

表 5 给出了二维联合模型中各个约束条件对中文触发词抽取的贡献,其中,在相似触发词事件联合学习中,触发词核心词素识别和触发词形态结构识别的准确率分别为 84.7%和 92.3%,这也确保了相似触发词事件联合学习的性能.

Table 5 Effectiveness of the different constraints in 2-dimension joint model (incremental constraints)

表 5 二维联合模型中不同约束条件的作用(约束条件递增)

系统	触发词识别			触发词类型分配		
	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i>	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i>
基准系统+论元语义	82.4	58.9	68.7	78.3	56.5	65.5
+TD 和 TTA 联合学习(约束 11、约束 12)	79.3	61.7	69.4	75.3	58.9	66.1
+相同触发词事件联合学习(约束 13)	79.5	63.5	70.6	75.4	60.6	67.2
+相似触发词事件联合学习(约束 14)	79.9	64.9	71.6	75.7	61.9	68.0
+相关事件联合学习(约束 15)	78.7	69.0	73.5	74.8	65.8	70.0
基准系统+论元语义+Li 等人 ^[1] 的联合模型	76.3	66.4	71.0	72.5	63.3	67.6

从表 5 的结果可知:

- 触发词识别 TD 和类型分配 TTA 的一维联合模型可分别提高两者的系统性能 0.7%和 0.6%,主要受益于召回率的提升(2.8%和 2.4%).主要原因是:由于 TTA 的结果既可帮助 TD 消除训练集中真假触发词实例失衡的不利影响,TTA 的结果也可帮助 TD 区分歧义信息.准确率下降是联合模型的一个通病,它往往会把一个假事件实例通过一致性推理推断成为真的事件实例,从而造成准确率下降.
- 相同触发词事件、相似触发词事件和相关事件联合学习可分别提高触发词识别的 *F1* 值 1.2%,1.0% 和 1.9%,验证了它们的有效性.其中,相同/相似触发词事件联合学习由于其本身具有较高一致性,所以其召回率和准确率均有提高.这主要由于在一个文档中,真相同/相似触发词实例之间存在高度一致性;同样,假的相同/相似触发词实例之间也存在高度一致性.利用它们进行联合学习,前者提高了系统的召回率,而后者则提高了准确率.相关事件联合学习大幅提高了系统的召回率,但是其准确率略有下降,这主要是由于相关事件的一致性概率相对而言较低,这就造成了部分假事件实例会被推断为真实例,从而影响准确率.总之,实验数据表明:利用关联事件语义进行事件触发词推理,根据已识别事件实例来推导与之关联的未识别事件实例是一种有效的方法.
- 本文提出的二维联合模型与目前性能最好的触发词识别和类型分配联合学习模型^[1]相比,分别提高了触发词识别和类型分配的 *F1* 值 2.5%和 2.4%.为了公平比较,该模型同样以加入了论元语义信息的触发词识别 TD 和类型分配 TTA 的结果为输入进行全局优化.这个结果充分说明,基于多种关联事件进行推理有助于提高中文触发词抽取的性能.

6 总结

本文针对中文触发词抽取中存在的中文事件的论元语义难以获取以及部分贫信息事件实例无法抽取的问题,提出了基于语义的中文事件触发词抽取联合模型.首先,根据中文句子结构灵活和句法成分多省略的特点,提出了基于模式匹配的核心论元和辅助论元抽取方法.这两类论元可以较好地表示论元语义,从而有机结合中文触发词抽取和论元抽取.其次,根据同一文档中关联事件实例间存在的高度一致性,构造了一个关联事件语义驱动的中文事件触发词识别和类型分配二维联合模型,用于抽取贫信息事件实例.实验结果表明:与现有的最好方法相比,本文提出的方法性能得到了明显的提升.

作为自然语言理解的重要基础,中文事件抽取是一个具有挑战性的任务,目前还处于起步阶段,有待进一步深入研究.我们下一步的研究工作将探讨更加有效的核心论元抽取以及表示方法,同时探索中文事件抽取 4 个环节的联合学习方法.

References:

- [1] 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.
- [2] Li PF, Zhu QM, Zhou GD. Argument inference from relevant event mentions in Chinese argument extraction. In: Proc. of the ACL 2013. Stroudsburg: ACL, 2013. 1477–1487.
- [3] Li Q, Ji H, Huang L. Joint event extraction via structural prediction with global features. In: Proc. of the ACL 2013. Stroudsburg: ACL, 2013. 73–82.
- [4] Kim YJ. Subject/Object drop in the acquisition of Korean: A cross-linguistic comparison. *Journal of East Asian Linguistics*, 2000, 9(4):325–351. [doi: 10.1023/A:1008304903779]
- [5] Riloff E. Automatically generating extraction patterns from untagged text. In: Proc. of the AAAI'96. Palo Alto: AAAI Press, 1996. 1044–1049.
- [6] Stevenson M, Greenwood M. A semantic approach to IE pattern induction. In: Proc. of the ACL 2005. Stroudsburg: ACL, 2005. 379–386. [doi: 10.3115/1219840.1219887]
- [7] Patwardhan S, Riloff E. Effective information extraction with semantic affinity patterns and relevant regions. In: Proc. of the EMNLP-coNLL 2007. Stroudsburg: ACL, 2007. 717–727.
- [8] 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.
- [9] Hardy H, Kanchakouskaya V, Strzalkowski T. Automatic event classification using surface text features. In: Proc. of the AAAI 2006 Workshop on Event Extraction and Synthesis. Palo Alto: AAAI Press, 2006. 36–41.
- [10] Finkel J, Grenager T, Manning C. Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proc. of the ACL 2005. Stroudsburg: ACL, 2005. 363–370. [doi: 10.3115/1219840.1219885]
- [11] 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.
- [12] 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.
- [13] Patwardhan S, Riloff E. A unified model of phrasal and sentential evidence for information extraction. In: Proc. of the EMNLP 2009. Stroudsburg: ACL, 2009. 151–160.
- [14] Lu W, Roth D. Automatic event extraction with structured preference modeling. In: Proc. of the ACL 2012. Stroudsburg: ACL, 2012. 835–844.
- [15] Ji H, Grishman R. Refining event extraction through cross-document inference. In: Proc. of the ACL 2008. Stroudsburg: ACL, 2008. 254–262.
- [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] 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]

- [18] Wang W. Chinese news event 5W1H semantic elements extraction for event ontology population. In: Proc. of the WWW 2012. New York: ACM Press, 2012. 197–202. [doi: 10.1145/2187980.2188008]
- [19] Chen C, Ng V. Joint modeling of Chinese event extraction with rich linguistic features. In: Proc. of the COLING 2012. Mumbai: Indian Institute of Technology Bombay, 2012. 529–544.
- [20] 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).
- [21] Fu JF, Liu ZD, Fu XF, Zhou W, Zhong ZM. Dependency parsing based event recognition. *Computer Science*, 2009,36(11):217–219 (in Chinese with English abstract).
- [22] Hou LB, Li PF, Zhu QM. Study of event recognition based on CRFs and cross-event. *Computer Engineering*, 2012,38(24):191–195 (in Chinese with English abstract).
- [23] Hou LB, Li PF, Zhu QM, Qian PD. Using cross-event inference to fill missing event argument. *Computer Science*, 2012,39(7): 200–204 (in Chinese with English abstract).
- [24] 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.
- [25] Chen Z, Ji H. Can one language bootstrap the other: A case study on event extraction. In: Proc. of the NAACL-HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing. Madison: Omnipress, 2009. 66–74.
- [26] Ji H. Cross-Lingual predicate cluster acquisition to improve bilingual event extraction by inductive learning. In: Proc. of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics (UMSLLS 2009). Madison: Omnipress, 2009. 27–35.
- [27] Qin B, Zhao YY, Ding X, Liu T, Zhai GF. Event type recognition based on trigger expansion. *Tsinghua Science and Technology*, 2010,15(3):251–258. [doi: 10.1016/S1007-0214(10)70058-4]
- [28] 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.
- [29] Zeng XD, Wong D, Chao L, Trancoso I. Graph-Based semi-supervised model for joint Chinese word segmentation and part-of-speech tagging. In: Proc. of the ACL 2013. Stroudsburg: ACL, 2013. 770–779.
- [30] Qian X, Liu Y. Joint Chinese word segmentation, POS tagging and parsing. In: Proc. of the EMNLP-CoNLL 2012. Stroudsburg: ACL, 2012. 501–511.
- [31] Li ZH, Zhang M, Che WX, Liu T. A separately passive-aggressive training algorithm for joint POS tagging and dependency parsing. In: Proc. of the COLING 2012. Mumbai: Indian Institute of Technology Bombay, 2012. 1681–1697.
- [32] Li JH, Zhou GD, Ng HT. Joint syntactic and semantic parsing of Chinese. In: Proc. of the ACL 2010. Stroudsburg: ACL, 2010. 1108–1117.
- [33] Riedel S, Chun HW, Takagi T, Tsujii J. A Markov logic approach to bio-molecular event extraction. In: Proc. of the BioNLP 2009. Stroudsburg: ACL, 2009. 41–49.
- [34] Poon H, Vanderwende L. Joint inference for knowledge extraction from biomedical literature. In: Proc. of the NAACL-HLT 2010. Madison: Omnipress, 2010. 813–821.
- [35] Riedel S, McCallum A. Fast and robust joint models for biomedical event extraction. In: Proc. of the EMNLP 2011. Stroudsburg: ACL, 2011. 1–12.
- [36] Do QX, Lu W, Roth D. Joint inference for event timeline construction. In: Proc. of the EMNLP-coNLL 2012. Stroudsburg: ACL, 2012. 677–687.
- [37] 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. Palo Alto: AAAI Press, 2013. 2120–2126.
- [38] Liao SS, Grishman R. Filtered ranking for bootstrapping in event extraction. In: Proc. of the COLING 2010. Beijing: Tsinghua University Press, 2010. 680–688.
- [39] Liu Q, Li SJ. Word similarity computing based on how-net. In: Proc. of the 3rd Chinese Lexical Semantic Workshop (CLSW 2002). 2002. (in Chinese with English abstract).
- [40] Grosz B, Aravind J, Scott W. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 1995,21(2):202–225.
- [41] Harris Z. *Mathematical Structures of Language*. New York: Wiley, 1968.

附中文参考文献:

- [20] 赵妍妍,秦兵,车万翔,刘挺.中文事件抽取技术研究.中文信息学报,2008,22(1):3-8.
- [21] 付剑锋,刘宗田,付雪峰,周文,仲兆满.基于依存分析的事件识别.计算机科学,2009,36(11):217-219.
- [22] 侯立斌,李培峰,朱巧明.基于 CRFs 和跨事件的事件识别研究.计算机工程,2012,38(24):191-195.
- [23] 侯立斌,李培峰,朱巧明,钱培德.基于跨事件理论的缺失事件角色填充研究.计算机科学,2012,39(7):200-204.
- [39] 刘群,李素建.基于《知网》的词汇语义相似度计算.见:第3届汉语词汇语义学研讨会会议论文.2002.

附录

Table 6 Core roles and their argument types

表6 事件核心角色及其核心论元类型

事件类型	核心角色(核心论元类型)
Merge-Org/End-Org/Start-Org/Declare-Bankruptcy	Org (ORG)
Transfer-Ownership/Transfer-Money	Buyer/Giver/Seller/Recipient (PER/ORG/GPE)
Transport	Agent (PER/ORG/GPE),Artifact (PER/VEH/WEA), Origin/Destination (GPE/LOC/FAC)
Attack	Attacker (PER/ORG/GPE),Target (PER/ORG/GPE/VEH/FAC/LOC)
Injure/Die	Agent (PER/ORG/GPE),Victim (PER)
Be-Born/Marry/Divorce	Person (PER)
Nominate/Execute/Arrest-Jail/Extradite	Person (PER),Agent (PER/ORG/GPE)
Start-Position/End-Position/Elect/Release-Parole	Person (PER),Entity (PER/ORG)
Demonstrate/Meet/Phone-Write	Entity (PER/ORG)
Convict/Acquit/Sentence/Trial-Hearing/ Charge-Indict/Appeal/Sue/Pardon	Defendant/Plaintiff/Adjudicator (PER/ORG/GPE)
Fine	Entity (PER/ORG),Adjudicator (PER/ORG/GPE)

Table 7 Supplement roles and their argument types

表7 辅助角色及其对应论元类型

事件类型	辅助角色(辅助论元类型)
Transfer-Ownership/Transfer-Money/Fine	Price/Money (MONEY)
Transport	Vehicle (VEH)
Attack/Injure/Die	Instrument (WEA/VEH/SUB)
Start-Position/End-Position/Elect/Nominate	Position (JOB)
Release-Parole/Execute/Arrest-Jail/Extradite/Convict/Acquit/ Sentence/Trial-Hearing/Charge-Indict/Appeal/Sue/Pardon/Fine	Crime (CRIME)
Sentence	Sentence (SEN)



李培峰(1971—),男,江苏苏州人,博士,副教授,CCF 高级会员,主要研究领域为自然语言处理,机器学习.



朱巧明(1963—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为自然语言处理,机器学习.



周国栋(1967—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为自然语言处理,机器学习.