

## 基于层叠条件随机场的事件因果关系抽取<sup>\*</sup>

付剑锋<sup>1,2</sup> 刘宗田<sup>2</sup> 刘 炜<sup>2</sup> 周 文<sup>2</sup>

<sup>1</sup>(上海立信会计学院 数学与信息学院 上海 201620)

<sup>2</sup>(上海大学 计算机工程与科学学院 上海 200027)

**摘 要** 传统的事件因果关系抽取方法只能覆盖文本中的部分显式因果关系. 针对这种不足, 提出一种基于层叠条件随机场模型的事件因果关系抽取方法. 该方法将事件因果关系的抽取问题转化为对事件序列的标注问题, 采用层叠(两层)条件随机场标注出事件之间的因果关系. 第一层条件随机场模型用于标注事件在因果关系中的语义角色, 标注结果传递给第二层条件随机场模型用于识别因果关系的边界. 实验表明, 本文方法不仅可以覆盖文本中的各类显式因果关系, 并且均能取得较好的抽取效果, 总体抽取效果的 F1 值达到 85.3%.

**关键词** 事件因果关系, 事件序列, 层叠条件随机场, 条件随机场模型  
中图法分类号 TP 391

### Event Causal Relation Extraction Based on Cascaded Conditional Random Fields

FU Jian-Feng<sup>1,2</sup>, LIU Zong-Tian<sup>2</sup>, LIU Wei<sup>2</sup>, ZHOU Wen<sup>2</sup>

<sup>1</sup>(School of Mathematics and Information, Shanghai Lixin University of Commerce, Shanghai 201620)

<sup>2</sup>(School of Computer Engineering & Science, Shanghai University, Shanghai 200072)

#### ABSTRACT

Traditional methods for event causal relation extraction covered only part of the explicit causal relation in the text. A method for event causal relation extraction is presented based on Cascaded Conditional Random Fields. The method casts the problem of event causal relation extraction as the labeling of event sequence. The Cascaded (Dual-layer) Conditional Random Fields is employed to label the causal relation of event sequence. The first layer of the Cascaded Conditional Random Fields model is used to label the semantic role of causal relation of the events, and then the output of the first layer is passed to the second layer for labeling the boundaries of the event causal relation. Experimental results show that this method not only covers each class of explicit event causal relation in the text, but also achieves good performance and the F-Measure of the overall performance arrives at 85.3%.

**Key Words** Event Causal Relation, Event Sequence, Cascaded Conditional Random Fields, CRFs

<sup>\*</sup> 国家自然科学基金(No. 60975033)、上海市重点学科开放课题项目(No. J50103)资助

收稿日期: 2010-02-08; 修回日期: 2010-11-25

作者简介: 付剑锋, 男, 1978 年生, 博士, 讲师, 主要研究方向为自然语言处理. E-mail: fujianfeng2000@gmail.com. 刘宗田, 男, 1946 年生, 教授, 博士生导师, 主要研究方向为语义智能、软件工程. 刘炜, 男, 1978 年生, 博士, 讲师, 主要研究方向为 Agent、事件本体. 周文, 女, 1979 年生, 博士, 讲师, 主要研究方向为概念格、事件本体.

## 1 引言

近年来,以事件作为知识表示和信息处理的单元受到越来越多的关注,各种基于事件的理论模型<sup>[1]</sup>和应用技术层出不穷<sup>[2-5]</sup>.人们通过研究事件的内部结构(时间、地点和参与者等事件要素)和外部关联(时序、因果等语义关系)来重新安排和组织文本(将自由文本转化为结构化文本),以支持基于事件的信息抽取<sup>[3]</sup>、问题回答<sup>[4]</sup>、自动文摘<sup>[5]</sup>等自然语言处理技术,进而实现对文本的理解.

在事件的外部关联中,因果关系(Causal Relation)是非常重要的一类语义关系,它反映了事件之间先后相继、由因及果的一种关系.研究事件的因果关系不仅可以用于预测和评估,在问题回答系统中也有着广泛的应用<sup>[4,6]</sup>.事件的因果关系主要由两部分构成:原因和结果.一般情况下,通常根据原因和结果之间的对应关系可以将因果关系分为一因一果、一因多果、多因一果和多因多果.实际上,根据原因和结果在文本中是否同时出现,我们首先可以将因果关系分为显式(Explicit)因果关系(原因和结果同时出现)和隐式(Implicit)因果关系<sup>[7]</sup>(在文本中省略了原因或者结果).隐式因果关系通常需要根据上下文知识进行推理,判断或者推测一个事件的原因或者结果,显式因果关系则不需要.因此,通常意义上的文本中的一因一果、一因多果等都属于显式因果关系的范畴.显式因果关系又可以分为带标记因果关系和无标记因果关系.带标记因果关系是指原因和结果之间由特定的语言成分连接,特定的语言成分可以是动词(如:导致,致使等)也可以是连词(如:因为…所以,由于等).反之,则是无标记因果关系.本文主要研究显式因果关系.

人们对事件因果关系的研究由来已久,早在亚里斯多德的《形而上学论》中就有过详细的阐述.在计算机领域, Joskowicz 等<sup>[8]</sup>和 Kaplan 等<sup>[9]</sup>针对特定领域的知识手工编写产生式规则(IF  $P$  THEN  $Q$ ,  $P \rightarrow Q$ )来发现隐式因果关系.国内干红华提出一种事件因果关系的结构分析方法<sup>[10]</sup>,并且采用产生式规则实现一个辅助法律咨询系统<sup>[11]</sup>.这种基于产生式规则的方法具有很强的领域特征,难以移植到其它领域.手工编写规则不仅费时费力,限制了它的大规模应用,还很难保证规则之间的兼容性(Compatibility)和完备性(Completeness).因此,近年来的研究转向利用自动或者半自动的方法从文本中抽取显式因果关系.

Garcia<sup>[12]</sup>分析了法语中表达因果关系的动词,并实现了一个 COATIS 系统用于抽取法语中形如 <Cause Verb Effect> 的带标记因果关系. Khoo 等<sup>[7]</sup>认为除了动词之外,还有些连词、形容词、副词以及子句也能表达因果关系,并采用模式匹配的方法从手工标注的 Wall Street Journal 语料中抽取带标记因果关系. Inui 等<sup>[13]</sup>采用机器学习的方法(SVM)从日文新闻中抽取显式带标记因果关系. Blanco 等<sup>[14]</sup>则采用了模式匹配和机器学习相结合的方法从 TREC5 语料上抽取带标记因果关系.类似工作还有文献[4]、[5]等.可以看出,上述方法都局限于抽取形如 <Cause Indicator Effect> 的带标记的因果关系,这种方法要求原因和结果之间存在特定连接成分,将其当作因果关系的指示词(Indicator).实际上,文本中存在大量的无特定连接成分的因果关系.如“曼哈顿唐人街一座六层楼廉价公寓 24 日凌晨发生大火,截止目前已有 2 名华人死亡,27 人受伤.”中的“大火→死亡”和“大火→受伤”.因此,Steven 等<sup>[15]</sup>开始尝试用机器学习的方法抽取显式因果关系(包括带标记和无标记因果关系).

值得注意的是,上述方法大多数都局限于抽取一对一的因果关系,无法处理一因多果的情况.因此,抽取的结果只能覆盖文本中的部分因果关系.此外,现有的研究方法主要抽取句内(Inner-Sentence)的因果关系,而在标注事件因果关系的过程中,我们发现文本中存在很多跨句、甚至是跨段的因果关系.针对这些不足,本文提出一种事件显式因果关系(本文简称因果关系)抽取方法——基于层叠条件随机场(Cascaded Conditional Random Fields, CCRFs)的事件因果关系抽取.与以往研究方法不同的是,该方法将因果关系的抽取问题转化为对事件序列的标注问题,并引入 CCRFs 模型对事件序列进行因果关系的语义角色标注和边界标注.这样不仅可以有效解决因果关系中的标记问题和对应关系问题,对于跨句、跨段因果关系也同样适用.实验表明,本文方法在各类因果关系上均可取得较好的抽取效果.

## 2 事件序列标注

在阐述本文方法中的事件因果关系抽取转化为事件序列标注之前,先介绍一些基本概念.

**定义 1(事件<sup>[1]</sup>(Event))** 在某个特定的时间和地点下发生的,由若干角色参与,表现出若干动作特征的一件事情.其中时间、地点、事件参与的对象称为事件要素.

**定义 2** (事件指示词 (Event Denoter))<sup>1)</sup> 文本中可以用来清晰地表示所发生的事情的词. 如地震、火灾、死亡、受伤、撞车等.

**定义 3** (事件序列 (Event Sequence)) 事件指示词按照它们在文本中先后出现的相对位置所组成的序列称之为事件序列. 如图 1 所示.

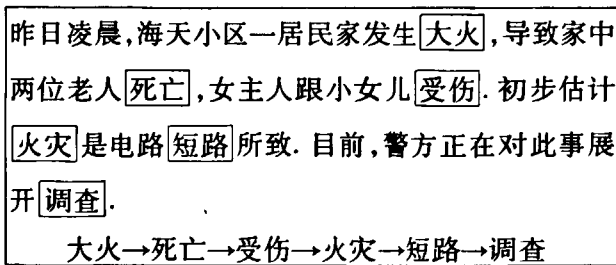


图 1 一个事件序列的例子

Fig. 1 An example of event sequence

将因果事件的抽取问题转化为事件序列的标注问题,就是通过对事件序列标注的方法,确定其中包含的因果关系. 标注过程分为两个步骤:1) 为事件序列中的每一个事件赋予标签,以表示其在因果关系中对应的语义角色;2) 为已经标注的因果关系贴上指界标签,划分因果关系的边界. 在 1) 中,我们引入一个标签集  $\psi = \{C, E, N\}$ , 其中  $C$  表示原因,  $E$  表示结果,  $N$  表示其它事件. 根据集合  $\psi$  中各个元素的定义,图 1 中事件序列的因果关系语义角色可标注为  $CEEECN$ . 对于 2),我们借鉴了组块分析 (Chunk Parsing) 中的  $BIO$  标签集. 其中,  $B$  表示因果关系的开始,  $I$  表示因果关系的中间或结尾,  $O$  为其它. 因此,图 1 中事件序列的因果关系边界可标注为  $BIIBIO$ . 具体过程如图 2 所示.

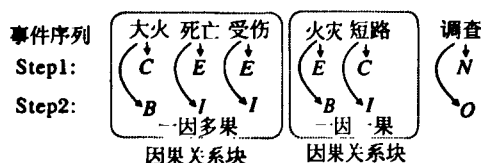


图 2 事件因果关系标注过程

Fig. 2 Process of event causal relation labeling

限于篇幅,本文没有列举出整篇文档的事件序列,但是可以理解,当文本中的因果关系在不同句子甚至是段落中时,采用事件序列标注的方法可有效应对这种跨句/段因果关系. 而且,该方法并不限于

一对一的因果关系,对于一因多果、多因一果以及多因多果的关系同样适用. 此外,在标注事件因果关系时,我们并不考虑原因和结果之间是否存在特定的连接成分,只要两者之间具有语义上的因果关系即被贴上相应的标签. 因此,本文方法也可以处理无标记因果关系.

对标注后的事件序列,只要采用一个反向解析的过程便可从中抽取出事件因果关系. 首先根据  $BIO$  标签抽取出因果系统块,然后再根据因果系统语义角色标签从因果系统块中抽取出原因和结果. 方法比较简单,不再赘述.

### 3 基于 CCRFs 的事件因果关系标注

#### 3.1 条件随机场模型

条件随机场<sup>[16]</sup> (Conditional Random Fields, CRFs) 是一种判别式的无向图模型,判别式模型与产生式模型不同,它不需要严格的独立性假设,并且可有效克服标注偏置等问题. 因此,被广泛应用于词性标注 (POS)、命名实体识别 (NER) 以及浅层句法分析 (Shallow Parsing) 等序列标注任务中. 研究表明,在单一模型的序列标注任务中,条件随机场模型可达到 state-of-the-art 的性能<sup>[17]</sup>.

线性链 (Linear-Chain) 是无向图的一种特例,由于序列标注的输出结果具有线性链的特性,因此线性链 CRFs 常用于为序列标注建模. 令  $X = x_1, x_2, \dots, x_n$  为可观测的输入序列 (比如: 词性标注任务中的句子、事件序列),  $Y = y_1, y_2, \dots, y_n$  为待预测的标记序列 (比如: 句子中单词所对应的词性、因果关系语义角色). 线性链 CRFs 定义标记序列  $Y$  的条件概率为

$$P_{\lambda}(Y|X) = \frac{1}{Z_X} \exp\left(\sum_{n=1}^N \sum_k \lambda_k f_k(y_{n-1}, y_n, X, n)\right),$$

$$Z_X = \sum_{y \in Y} \exp\left(\sum_{n=1}^N \sum_k \lambda_k f_k(y_{n-1}, y_n, X, n)\right),$$

其中,  $Z_X$  是归一化因子,  $f_k(y_{n-1}, y_n, X, n)$  是特征函数 (对于离散型特征,函数的取值通常为  $\{0, 1\}$ ),  $\lambda_k$  是特征  $f_k$  对应的权重. 特征函数  $f_k(y_{n-1}, y_n, X, n)$  可计算在当前时刻  $n$  时标记转移  $y_{n-1} \rightarrow y_n$  以及输入序列  $X$  的任意特征的值. 比如: 当  $x_{n-1}$  为事件指示词“地震”,  $x_n$  为事件指示词“死亡”,  $y_{n-1}$  为标记  $C$ ,  $y_n$  为标记  $E$  时,特征函数可能取值为 1. 因此,可以将任意的、具有依赖关系特征加入条件随机场模型中,

1) 在某些参考文献中,事件指示词也称为事件触发词 (Event Trigger).

而不用像产生式模型一样必须单独计算它们的分布. 条件随机场模型的时间复杂度为  $O(NFS^2)$ , 其中  $N$  为序列的长度,  $F$  为特征数,  $S$  为标记的个数.

特征函数的权重可通过模型训练获得相应的估值, 通常使用最大似然估计法. 其似然函数为

$$L_A = \sum_{i=1}^M \log(P_A(Y_i | X_i)) - \sum_k \frac{\lambda_k^2}{2\sigma_k^2},$$

上式等号右边第 2 项是高斯先验值, 为了避免训练时出现过拟合 (Over-Fitting) 加入的惩罚因子, 其中  $\sigma^2$  是先验方差. 似然函数  $L_A$  是一个凸函数, 可以保证条件随机场模型在训练时经过不断的迭代, 目标函数收敛在全局最优值. 本文采用 L-BFGS (Limited Memory BFGS)<sup>[18]</sup> 算法来求解似然函数的最优值.

对条件随机场模型的解码即, 给定一个输入序列  $X$ , 寻找以输入序列  $X$  为条件下具有最大条件概率的标记序列

$$Y^* = \arg \max_y P_A(Y | X),$$

解码过程一般采用动态规划算法, 本文采用 Viterbi 算法.

### 3.2 事件因果关系标注的 CCRFs 模型

虽然条件随机场模型在序列标注中有非常好的性能, 但是实际应用当中有些问题比较复杂 (比如在中文组块识别中, 既要识别出组块的边界, 又要识别组块的类型), 采用单层条件随机场模型往往达不到最佳效果. 因此, 需要引入多层 CRFs 模型来解决这些问题<sup>[19]</sup>. 其基本思想是: 先将复杂的任务分解为几个子任务, 再对每一个子任务使用条件随机场模型. 本文将事件因果关系的标注过程分为了两个步骤: 事件因果关系语义角色标注; 事件因果关系边界标注. 我们采用按层叠加的方法为其建立一个两层条件随机场模型, 每 1 层 CRFs 对应 1 个子任务. 在训练阶段, 分别的对第 1 层和第 2 层 CRFs 进

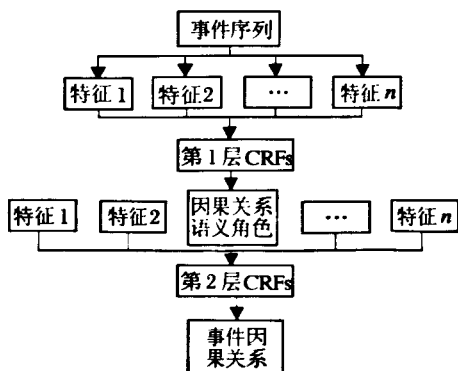


图3 事件因果关系标注的 CCRFs 模型

Fig.3 CCRFs model of event causal relation labeling

行训练. 其中, 第 2 层 CRFs 模型的输入特征还包括事件因果关系的语义角色特征. 在解码阶段, 第 1 层模型用于事件因果关系的语义角色标注, 标注结果作为特征传递给第 2 层模型, 用于辅助事件因果关系边界的标注. 如图 3 所示.

在事件因果关系标注的 CCRFs 模型中, 第 1 层 CRFs 和第 2 层 CRFs 之间是一种线性组合关系. 模型的时间复杂度为  $O(LNFS^2)$ , 其中  $L$  为层数.

### 3.3 特征选择和归纳

特征选择 (Feature Selection) 是运用条件随机场模型于实际应用中的一个重要环节, 需要根据特定的任务来选择与之相适应的特征. 条件随机场模型允许选择任意的特征, 因此可以针对特定的任务在特征中加入相关的领域知识 (比如: 英文命名实体识别的领域知识有词性、首字母是否大写等等), 这些领域知识也称为领域特征 (Domain Features). 在事件因果关系语义角色标注的任务中 (即第 1 层 CRFs), 事件的指示词、事件所属的类型以及事件的极性 (肯定或者否定) 等都是非常重要的领域特征. 以事件的极性为例, “火灾造成 1 人死亡, 2 人重伤.” 是一个一因多果关系 (火灾 → 死亡, 火灾 → 重伤), 而 “此次火灾没有造成人员伤亡.” 中的 “火灾” 和 “伤亡” 之间则不存在因果关系. 在本文中, 针对事件因果关系语义角色标注选择的领域特征如表 1 所示.

表 1 事件因果关系语义角色标注选择的领域特征

Table 1 Domain features for semantic role labeling of event causal relation

No.	领域特征	特征的意义
1	Denoter	事件指示词
2	DenoterPOS	事件指示词的词性
3	DenoterPosition	指示词的相对位置 <sup>1)</sup> : 文章首位、文章末尾、段落首位、段落末尾和其它
4	EventType	事件所属的类型: 包括 emergency、state changing、operation 等等
5	EventPolarity	事件的极性 (肯定或否定)

表 1 中的领域特征还应在窗口  $\pm 3$  的范围内进行偏移, 以容纳更多的上下文信息. 对于本文中的层叠条件随机场模型中的第 2 层——事件因果关系边界标注, 则还应加上事件因果关系的语义角色作为领域特征.

自然语言具有复杂多变、前后关联的特点, 仅用

1) 指示词的相对位置是指将文本用事件序列表示后的指示词之间的位置关系. 比如: “文章首位” 是指事件序列中出现的第一个事件指示词, 并非出现在文章中的第一个词.

上述领域特征还不足以描述事件序列中事件的复杂性以及事件之间的关联关系. 而条件随机场模型具有表达长距离依赖和交叠特性的能力, 因此还需要构造一些关联特征 (Conjunction Features) 来表达这些复杂关系. 关联特征是通过领域特征之间的某种组合关系来实现的, 比如: 前后两个事件指示词的组合, 下一个事件的指示词、事件所属类型以及事件极性的组合等. 但是, 特征之间的组合将导致特征集的迅速膨胀, 而且特征集中还可能存在一些“噪声”特征, 这些都会影响模型训练和解码的性能. 因此需要从特征集中去掉“噪声”特征, 优选出一个对标注任务贡献较大的特征子集. 此外, 我们还希望能自动的实现特征组合, 尽量避免人工的干预. 这种自动组合特征并且进行特征选择的过程, 称为特征归纳<sup>[20]</sup> (Feature Induction).

本文采用一种基于信息增益 (Information Gain) 的特征归纳方法, 该方法主要通过计算特征的增益, 然后根据一定的规则构造出一个候选特征集, 再从候选特征集选出最优的特征子集. 具体算法步骤如下.

step 1 将最优特征子集置空.

step 2 将领域特征加入候选特征集, 评估每一个领域特征的增益.

step 3 选择具有高增益的领域特征 (Top 5) 进行组合, 加入候选特征集.

step 4 对候选特征集中的所有特征进行评估, 将高增益的特征 (Top 10) 加入最优特征子集.

step 5 采用 L-BFGS 算法重新计算最优特征子集中每一个特征的权重.

step 6 重复 step 2 ~ step 5, 直到收敛.

上述算法中, 我们借鉴文献[20]中的评估函数计算特征  $f$  的增益, 如下所示:

$$G_A(f, \lambda) = \sum_{i=1}^N \log \left( \frac{\exp(\lambda f(y_{i(i)}, X_{(i)}, n(i)))}{Z_{X(i)}(\lambda f, \lambda)} \right) - \frac{\lambda^2}{2\sigma^2}.$$

采用该算法在实验语料 (具体可参见本文实验部分) 上针对事件因果关系标注中的每 1 层 CRFs 分别进行了特征归纳. 第 1 层 CRFs 得到 1 780 个特征, 第 2 层 CRFs 得到 2 030 个特征. 限于篇幅, 我们仅列举出其中的 8 个特征样例, 详细情况如表 2 所示. 其中, 特征 1-5 为第 1 层 CRFs 和第 2 层 CRFs 所共有, 特征 6-8 仅为第 2 层 CRFs 所独有.

表 2 特征归纳得到的部分特征样例

Table 2 A Sample of some features result from feature induction

No.	特征	特征说明
1	EventType(n) = "Emergency"	当前事件类型为 "Emergency"
2	Denoter(n) = "火灾" & Denoter(n+1) = "短路"	当前事件指示词为 "火灾" 并且右边第一个事件指示词为 "短路"
3	Denoter(n) = "爆炸" & EventType(n) = "Emergency"	当前事件指示词为 "爆炸" 并且事件类型为 "Emergency"
4	SemanticRole(n) = "N"	当前事件的因果关系语义角色为 "N"
5	Denoter(n) = "火灾" & SemanticRole(n) = "C" & DenoterPosition(n) = "BD"	当前事件指示词为 "火灾" 并且对应的因果关系语义角色为 "C" 并且该指示词位于 "BD" (文章首位)

## 4 实验

### 4.1 语料准备和分析

采用机器学习的方法抽取事件因果关系需要相应的语料库来进行训练和测试. 虽然对于事件的信息抽取在消息理解会议 (Message Understanding Conference, MUC) 和自动内容抽取会议 (Automatic Content Extraction, ACE) 中都有相关的评测, 并且提供了相应的评测语料. 但是上述会议主要关注于事件模板的填充或者事件要素的识别. 目前, 国内外对于事件因果关系的研究才刚刚起步, 尚未发现有可供公开评测的语料库. 已有的部分研究工作<sup>[7,15]</sup> 主要

采用了手工的方法标注事件的因果关系. 然而, 正如本文第 1 节所述, 他们的方法 (包括语料) 局限于句内的一因一果关系. 在我们的前期研究工作<sup>[21]</sup> 中, 从网上收集了 200 篇关于地震、火灾、交通事故、食物中毒以及恐怖袭击等 5 类突发事件的新闻报道作为生语料, 并且对语料中的事件进行了标注. 在此基础上, 我们进一步标注了事件之间的因果关系.

具体步骤如下.

step 1 将标注人员分为两组, 分别独立的对语料进行标注, 在标注过程中只允许组内讨论.

step 2 对两组标注后的语料进行一致性检查, 标注一致的因果关系, 则认为标注正确.

step 3 对于标注不一致的地方, 邀请第 3 方参

与进来一起进行投票表决。

我们对上述 200 篇语料进行了标注,共发现 340 个因果关系,详细情况如表 3 所示。

表 3 语料中的事件因果关系分布情况

Table 3 Distribution of event causal relation of corpus

分类方法	关系类型	数量	所占比例/%
原因和结果的对应关系	一因一果关系	166	48.8
	一因多果关系	173	50.9
	多因一果关系	1	0.3
	多因多果关系	0	0.0
是否包含特定连接成分	带标记因果关系	132	38.8
	无标记因果关系	208	61.2
在文本中的位置关系	句内因果关系	248	72.9
	跨句因果关系	50	14.7
	跨段因果关系	42	12.4

从表 3 中可以看出,带标记、句内的一因一果关系实际上是 3 种不同分类标准的一个交集,因此,传统方法只能覆盖文本中的部分因果关系.虽然因果关系按照对应关系可分为一因一果、一因多果、多因一果和多因多果,但是从表 3 中可以看出,现实世界对于突发事件的新闻报道中包含的多因一果和多因多果关系非常少.考虑到语料稀疏的问题,我们在后续实验中暂不考虑这两种关系.但是,理论上本文方法对这两种关系的抽取是同样有效的。

4.2 实验结果和分析

为了验证本文方法在事件因果关系上的抽取效果,我们在上述语料上进行了实验.实验采用随机的方法抽取语料的 3/4 作为训练集,剩下的 1/4 作为测试集.为了避免因语料不平衡带来的影响,我们将实验重复 10 次,结果取均值.实验结果采用精确率(Precision)、召回率(Recall)和 F1 值(F-Measure)这 3 个指标来评价.由于事件的因果关系涉及到原因和结果两个方面,因此,我们在判断一个因果关系是否被正确抽取时,要求原因和结果同时正确(如果是多个结果,则要求多个结果同时正确),否则视为抽取错误.实验结果如表 4 所示。

从表 4 可以看出,一因一果关系的抽取效果要好于一因多果关系,因为一因多果关系有多个结果需要抽取(语料中的结果最多可达 6 个),这是造成抽取效果不如一因一果关系的主要原因.在选择特征的时候,我们并未考虑特定连接成分这一因素,所选特征没有特别偏向于带标记因果关系,因此带标记因果关系与无标记因果关系的抽取效果基本相当.从另外一个角度来讲,本文方法对无标记因果关系也是同样有效的.对于跨句和跨段因果关系,由于

原因和结果之间的文本距离较远,中间存在噪音的干扰(比如非因果关系中的事件),导致抽取效果不如句内因果关系.总体上,文本方法对因果关系的抽取达到了 85.3% 的 F1 值。

表 4 事件因果关系抽取结果

Table 4 Experimental results for event causal relation extraction

关系类型	precision	recall	F1
一因一果关系	95.3%	85.7%	90.2%
一因多果关系	83.9%	78.3%	81.0%
带标记因果关系	90.6%	81.4%	85.8%
无标记因果关系	88.3%	81.9%	85.0%
句内因果关系	91.8%	83.6%	87.5%
跨句因果关系	84.7%	77.6%	81.0%
跨段因果关系	79.2%	75.5%	77.3%
Overall	89.2%	81.7%	85.3%

此外,为了量化本文方法中特征归纳对于系统性能的改善效果,我们还进行 1 组对比实验,将没有采用特征归纳(仅使用领域特征和上下文特征)与采用特征归纳的事件因果关系抽取效果进行对比.实验结果如表 5 所示。

从表 5 可以看出,在没有采用特征归纳的情况下,系统只取得 63.9% 的 F1 值,而采用了特征归纳之后,系统的 F1 值达到了 85.3%,提高了 21.4%.因此,特征归纳可以显著改善系统性能,提高事件因果关系的抽取效果。

与以往研究相比,由于实验语料不相同,而且所能覆盖到的因果关系范围也不相同,所以本文不便直接比较实验结果,我们仅列举出 2 个数据以供参考. Blanco 等<sup>[14]</sup>采用了模式匹配和机器学习相结合的方法来抽取带标记因果关系,手工标注了 SemCor2.1 语料,共有 517 句子包含因果关系,抽取效果达到了 89.5% 的 F1 值. Steven 等<sup>[15]</sup>在 TreeBank 语料上手工标注了 271 个因果关系(一因一果),采用机器学习的方法抽取因果关系,不过只取得 52.4% 的 F1 值。

表 5 系统性能比较

Table 5 Comparison of system performance

	precision	recall	F1
Without feature induction	68.1%	60.2%	63.9%
With feature induction	89.2%	81.7%	85.3%

5 结束语

事件的因果关系是一类语义关系,

研究事件因果关系在事件预测、评估以及问题回答等领域有广阔的应用前景. 现有研究只能覆盖文本中的部分因果关系, 针对这种不足, 本文提出了一种基于层叠条件随机场的事件因果关系抽取方法. 该方法将事件因果关系的抽取问题转化为对事件序列的标注问题, 采用层叠(两层)条件随机场标注出事件之间的因果关系. 第1层条件随机场模型用于标注事件在因果关系中的语义角色, 标注结果传递给第2层条件随机场模型用于识别因果关系的边界. 为了提高系统性能, 本文还设计了一种基于信息增益的特征归纳方法, 实现了特征的自动组合, 并从中选择最优的特征子集, 实验表明, 特征归纳可以显著提高事件因果关系的抽取效果. 我们从网上收集了对现实世界中突发事件的新闻报道, 标注了其中的事件之间的因果关系. 从语料分析可以看出, 该方法可以有效覆盖文本中的因果关系. 实验表明, 本文方法在各类因果关系上均能取得较好抽取效果, 总体抽取效果较为理想.

**致 谢** 感谢朱莎莎、朱芳、徐文杰、蔡毅以及上海大学语义智能实验室的其他成员为语料标注做了大量的工作.

### 参 考 文 献

- [1] Liu Zongtian, Huang Meili, Zhou Wen, *et al.* Research on Event-oriented Ontology Model. *Computer Science*, 2009, 36(11): 189 – 192 (in Chinese)  
(刘宗田, 黄美丽, 周文, 等. 面向事件的本体研究. *计算机科学*, 2009, 36(11): 189 – 192)
- [2] Ahn D. The Stages of Event Extraction // *Proc of the COLING-ACL Workshop on Annotating and Reasoning about Time and Events*. Sydney, Australia, 2006: 1 – 8
- [3] Surdeanu M, Harabagiu S, Williams J, *et al.* Using Predicate-Argument Structures for Information Extraction // *Proc of the 41st Annual Meeting on Association for Computational Linguistics*. Sapporo, Japan, 2003: 8 – 15
- [4] Girju R. Automatic Detection of Causal Relations for Question Answering // *Proc of the 41st ACL Workshop on Multilingual Summarization and Question Answering*. Sapporo, Japan, 2003: 76 – 83
- [5] Filatova E, Hatzivassiloglou V. Event-Based Extractive Summarization // *Proc of ACL Workshop on Summarization*. Barcelona, Spain, 2004: 104 – 111
- [6] Pechsiri C, Kawtrakul A. Mining Causality from Texts for Question Answering System. *IEICE Trans on Information and Systems*, 2007, 90(10): 1523 – 1533
- [7] Khoo C, Kornfilt J, Oddy R, *et al.* Automatic Extraction of Cause-Effect Information from Newspaper Text without Knowledge-Based Inferencing. *Literary and Linguistic Computing*, 1998, 13(4): 177 – 186
- [8] Joskowicz L, Kasiezyc T, Grishman R. Deep Domain Models for Discourse Analysis // *Proc of the Annual AI Systems in Government Conference*. Washington D C, USA, 1989: 195 – 200
- [9] Kaplan R, Berry-Rogghe G. Knowledge-Based Acquisition of Causal Relationships in Text. *Knowledge Acquisition*, 1991, 3(3): 317 – 337
- [10] Gan Honghua, Pan Yunhe. A New Analysis of the Structure of Event Causation. *Pattern Recognition and Artificial Intelligence*, 2003, 16(1): 56 – 62 (in Chinese)  
(干红华, 潘云鹤. 一种基于事件的因果关系的结构分析方法. *模式识别与人工智能*, 2003, 16(1): 56 – 62)
- [11] Gan Honghua. A Computational Analysis of Event Causation: Theory and Applications. Ph. D Dissertation. Hangzhou, China: Zhejiang University. School of Computer Science, 2003 (in Chinese)  
(干红华. 基于事件的因果关系可计算化分析研究. 博士论文. 杭州, 中国: 浙江大学. 计算机科学与技术学院, 2003)
- [12] Garcia D. COATIS, an NLP System to Locate Expressions of Actions Connected by Causality Links // *Proc of the 10th European Workshop on Knowledge Acquisition, Modeling and Management*. Catalonia, Spain, 1997: 347 – 352
- [13] Inui T, Inui K, Matsumoto Y. What Kinds and Amounts of Causal Knowledge Can Be Acquired from Text by Using Connective Markers as Clues? *Lecture Notes in Artificial Intelligence*, 2003: 180 – 193
- [14] Blanco E, Castell N, Moldovan D. Causal Relation Extraction // *Proc of the 6th International Conference on Language Resources and Evaluation*. Marrakech, Morocco, 2008: 310 – 313
- [15] Bethard S, Martin J H. Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations // *Proc of the 46th Annual Meeting of ACL on Human Language Technologies*. Columbus, USA, 2008: 177 – 180
- [16] Lafferty J, McCallum A, Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data // *Proc of the 18th International Conference on Machine Learning*. Williamstown, USA, 2001: 282 – 289
- [17] Sha F, Pereira F. Shallow Parsing with Conditional Random Fields // *Proc of the Conference of the North America Chapter of the Association for Computational Linguistics on Human Language*. Edmonton, Canada, 2003: 213 – 220
- [18] Nocedal J, Wright S. Numerical Optimization. New York, USA: Springer-Verlag, 1999
- [19] Sutton C, McCallum A. Composition of Conditional Random Fields for Transfer Learning // *Proc of the Human Language Technology Conference on Empirical Methods in Natural Language Processing*. Vancouver, Canada, 2005: 748 – 754
- [20] McCallum A. Efficiently Inducing Features of Conditional Random Fields // *Proc of the 19th Conference on Uncertainty in Artificial Intelligence*. Acapulco, Mexico, 2003: 2 – 35
- [21] Fu Jianfeng, Liu Wei, Liu Zongtian. A Study of Chinese Event Taggability // *Proc of the 2nd International Conference on Communication Software and Networks*. Singapore, Singapore, 2010: 400 – 404