

# 基于 MLN 的中文事件触发词推理方法

朱少华<sup>1,2</sup> 李培峰<sup>1,2,†</sup> 朱巧明<sup>1,2</sup>

1. 苏州大学自然语言处理实验室, 苏州 215006; 2. 苏州大学计算机科学与技术学院, 苏州 215006;

† 通信作者, E-mail: pfli@suda.edu.cn

**摘要** 现有的中文事件触发词抽取方法大多数采用特征工程和触发词扩展方法, 无法利用同一文档中各个触发词实例之间的内在关系。为了解决上述问题, 基于马尔科夫逻辑网络(MLN), 利用核心词素, 训练语料中触发词实例填充真假事件的概率, 以及触发词实例间的关系等信息来推导测试集中缺乏有效上下文信息和低可信度的触发词实例。在 ACE 2005 中文语料上的实验结果表明, 与基准系统相比, 该方法在触发词识别和事件类型分类阶段 F1 值分别提高 3.65% 和 2.51%。

**关键词** 触发词抽取; 马尔科夫逻辑网络; 触发词推理

**中图分类号** TP391

## A Chinese Event Trigger Inference Approach Based on Markov Logic Networks

ZHU Shaohua<sup>1,2</sup>, LI Peifeng<sup>1,2,†</sup>, ZHU Qiaoming<sup>1,2</sup>

1. Natural Language Processing Lab, Soochow University, Suzhou 215006; 2. School of Computer Science and Technology, Soochow University, Suzhou 215006; † Corresponding author, E-mail: pfli@suda.edu.cn

**Abstract** Previous Chinese argument extraction approaches mainly focus on feature engineering and trigger expansion, which cannot exploit inner relation between trigger mentions in same document. To address this issue, the authors bring forward a novel trigger inference mechanism based on Markov logic network. Head morpheme, the probabilities of a trigger mention fulfilling true and pseudo events from the training set and the relationships between trigger mentions are used to infer those trigger mentions with lack of effective context information or low confidences in testing set. Experimental results on the ACE 2005 Chinese corpus show that the proposed approach outperforms the baseline, with the F1 improvements of 3.65% and 2.51% in trigger identification and event type classification respectively.

**Key words** trigger extraction; Markov logic network; trigger inference

作为信息抽取的一个关键部分, 事件抽取任务是识别预定义的事件类型的实例及其论元(参与者和属性), 包括两个子任务: 1) 触发词抽取, 用于探测事件实例及其类型; 2) 论元抽取, 用于探测事件的论元及其角色。触发词抽取又分为两部分: 1) 触发词识别, 用于识别事件的触发词; 2) 事件类型分类, 赋予每个抽取触发词实例一个事件类型。本文工作聚焦在触发词抽取方面。

为了更好地理解 ACE 评估中的中文事件抽取任务, 下面列出本文所涉及的部分概念。

**事件(Event):** 在真实世界中已经、可能或将要发生的事情, 一般包括时间、地点和人物等角色, 如出生、地震和车祸等事件。

**事件实例(Event Mention):** 在文档中描述一个事件的句子或子句。

**事件类型(Event Type):** 事件的类别, 如 Attack

(攻击)和 Injure (受伤)等。

触发词(Trigger): 用于标识事件的谓词(又称为锚), 是事件的基本要素之一。如“生于”、“出生”就是 Be-Born (出生)事件的触发词。事件识别关键就是识别事件的触发词。

触发词实例(Trigger Mention): 触发词在句子中的提及。

论元(Argument): 与相关事件的实体实例, 是构成事件的基本要素之一。

角色(Role): 表述论元和事件的关系, 如出生事件中的角色有人物、出生地点、出生时间等。

E1 是一个事件实例, 事件类型是 Transport (运送), 包括触发词“逃离”及相应论元“米洛舍维奇”和“贝尔格勒”, 两个论元在该事件实例中担任的角色分别是“Artifact”和“Origin”。本文所有例子全部取自 ACE2005 中文语料库<sup>[1]</sup>。

E1: 米洛舍维奇 (Artifact) 被迫逃离 (Movement: Transport) 贝尔格勒 (Origin)。

现有的中文触发词抽取大多采用特征工程方法, 通常采用句子级别的特征, 并且假设各个实例相互独立, 无法利用同一文档中各个触发词实例之间的一致性关系。本文基于马尔科夫逻辑网络, 利用触发词或核心词素同指一致性和相关一致性等推理规则, 学习训练语料中候选触发词填充真假事件的概率和触发词实例间的关系, 从而进行中文触发词推理。

## 1 相关工作

绝大多数的事件抽取研究聚焦在英文。早期的研究侧重在句子级别抽取<sup>[2-4]</sup>, 后来转向运用更高级别的信息, 比如文档<sup>[5-6]</sup>、跨文档<sup>[7]</sup>、跨事件<sup>[8]</sup>和跨实体<sup>[9]</sup>信息。

相比于在英文触发词抽取上的大量研究, 只有少量研究是关于中文触发词抽取的。从研究方法上而言, 中文事件抽取主要从特征选择和触发词扩展两个方面入手。

在特征选择方面, Tan 等<sup>[10]</sup>提出一种局部特征选择方法, 并运用多层模板来提高中文事件抽取系统的性能; Chen 等<sup>[11]</sup>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特征, 用于中文事件抽取; Fu 等<sup>[12]</sup>提出一个特征加权方法, 用于对各种特征进行权重的重

新分配, 并用于触发词的识别和事件分类; Wang<sup>[13]</sup>把各种特征(如项频度、句子位置和长度、标题词覆盖率、语义角色标注等)组合起来, 选择信息化最大的句子作为事件的候选; Chen 等<sup>[14]</sup>把字符信息、语义角色标注信息、触发词概率信息、零指代信息、触发词一致性信息和论元一致性信息等特征用于中文事件抽取。

在触发词扩展方面, Chen 等<sup>[15]</sup>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, 尝试从跨语言事件抽取中利用联合训练来提高中文和英文事件抽取性能; Ji<sup>[16]</sup>从中英平行语料库入手, 在英文中利用中英翻译来扩展中文触发词; Qin 等<sup>[17]</sup>则用“同义词词林”来扩展中文事件触发词; Li 等<sup>[18-19]</sup>根据中文词组的组合语义学原理, 分别从动词构词结构和形态结构两个方面入手, 识别在训练语料中不出现的未知触发词, 并结合篇章级别的一致性进行事件识别。

除特征选择和触发词扩展两种方法外, 部分研究采用推理方法进行触发词抽取。Chklovski 等<sup>[20]</sup>利用词-句匹配模板抽取具有事件关系的资源, 并将抽取的结果整理成一个称为“VerbOcean”的知识库; Pantel 等<sup>[21]</sup>通过 Espresso 算法进行自动模板的构建, 在给定少量关系实例的情况下, 通过机器学习方法进行迭代扩展, 最终得到大量的关系模板。

## 2 中文触发词推理

基于特征工程的方法假定各个实例之间相互独立, 无法利用实例之间的内在联系。马尔科夫逻辑可以利用实例之间的内在一致性进行触发词推理。

### 2.1 马尔科夫逻辑网络

马尔科夫逻辑网络(Markov Logic Network, MLN)<sup>[22]</sup>是一门基于一阶逻辑和马尔科夫网络的统计关系学习语言。一个一阶知识库可以视为在可能域上的强约束集合: 如果一个域违反即使一条规则, 则该规则概率为0。MLN 的基本思想是软化这些约束: 当一个域违反知识库中的一条规则时, 概率会变小但是并非不可能。违反规则的域越少, 这条规则成立的可能性就越大。

一个 MLN 可以看做构建马尔科夫网络的模板。给定不同的常量集合, 会产生不同的网络。在这样一个网络中, 状态  $x_i$  的概率为

$$P(x) = (1/Z) \exp\left(\sum_i w_i f_i(x)\right),$$

其中,  $Z$  为归一化因子; 每一条规则下的不同常量组合对应一条子句,  $w_i$  为第  $i$  条子句的权值,  $f_i(x)$  为第  $i$  条子句的特征, 如果第  $i$  条子句为真, 则  $f_i=1$ , 否则  $f_i=0$ 。

MLN 已经成功应用到自然语言处理中, 如信息抽取、指代消解、语义解析、实体解析, 取得不错的效果。

MLN 推理主要包括谓词说明和推理公式, 表 1 为本文定义的谓词。另外, 本文只有一个查询谓词  $\text{Event}(\text{tri}_i, \text{label}_i, \text{doc}_i)$ , 即触发词  $\text{tri}_i$  在  $\text{doc}_i$  文档中对应的事件类型为  $\text{label}_i$  (0 或 1)。

本文利用事件类型一致性(规则(1))、触发词同指一致性(规则(2))、核心词素同指一致性(规则(3))、相关事件一致性(规则(4))等触发词推理规则对触发词进行推理。

## 2.2 事件类型一致性推理

触发词抽取的两个模块在系统级别也存在一致性。本文利用触发词识别和事件类型分类的一致性来恢复真触发词。也就是说, 如果一个触发词实例被事件类型分类模块识别为一个特定类型的事件, 那么它必然被触发词识别模块识别为真触发词。

**事件类型一致性推理(规则(1))** 如果一个候选触发词  $\text{tri}_i$  在文档  $\text{doc}_i$  中, 且在分类器中分类为某一类事件  $\text{evType}_i$  的概率很高, 而在触发词识别中被分类为假触发词(未触发事件), 则可以认为该候选触发词  $\text{tri}_i$  有很高的概率为真触发词( $\text{label}_i$  推理为 1)。

$$\text{EvType}(\text{tri}_i, +\text{evType}_i, \text{doc}_i) \Rightarrow \text{Event}(\text{tri}_i, \text{label}_i, \text{doc}_i), \quad (1)$$

其中, “+”表示针对每个事件类型, MLN 都要从训练集中学习出一个独立的权值。

## 2.3 触发词同指一致性推理

汉语作为话题结构的语言, 在同一文档中, 相

同触发词触发的事件大多数情况下相同<sup>[23]</sup>(同指事件)。现有的中文触发词抽取大多采用特征工程方法。该方法假设各个实例相互独立, 无法利用同一文档中相同触发词实例之间的同指一致性关系。以 E2 和 E3 为例。

E2: 美国与北韩 3 号在吉隆坡结束飞弹会议(EM1)。

E3: 会谈(EM2)的气氛严肃。

E2 和 E3 出现在同一文档中。很明显, E2 中的触发词实例“会谈”由于其上下文存在时间、地点和会谈者等角色, 相对较容易被识别为 Meet (会见)事件; E3 中的相同触发词实例“会谈”却由于缺乏有效的上下文信息而往往难以被识别。在同一文档内提及的触发词实例往往有语义上的一致性: 如果该触发词的其中一个实例触发了一个事件, 那么该文档内其他的实例有很高概率触发同类事件。所以, 可以从富信息的事件实例推断出贫信息的事件实例。

**触发词同指一致性推理(规则(2))** 候选触发词  $\text{tri}_i$  和  $\text{tri}_j$  相同, 且它们在同一文档  $\text{doc}_i$  中, 如果  $\text{tri}_i$  触发了事件, 那么  $\text{tri}_j$  也触发事件。

$$\text{SameTri}(\text{tri}_i, \text{tri}_j, \text{doc}_i) \wedge \text{Event}(\text{tri}_i, +\text{label}_i, \text{doc}_i) \Rightarrow \text{Event}(\text{tri}_j, +\text{label}_j, \text{doc}_i). \quad (2)$$

## 2.4 核心词素同指一致性推理

形态结构相似的触发词也具有话题内一致性。在 E4 中存在 3 个事件实例(EM1, EM2, EM3), 3 个触发词实例“担任”、“出任”和“升任”虽然不相同, 但是它们具有相同的核心词素“任”以及相同的形态结构。核心词素是在中文动词或者名词中占据主导语义的元素, 例如触发词“新婚”的核心词素为“婚”。这 3 个触发词都指向相同的事件类型 Start-Position (任新职), 它们具有语义上的一致性。在同一文档内提及的具有相同核心词素且触发词形态结

表 1 谓词及其说明  
Table 1 Predicates and descriptions

谓词	说明
$\text{EvType}(\text{tri}_i, \text{evType}_i, \text{doc}_i)$	触发词 $\text{tri}_i$ 对应的事件类型为 $\text{evType}_i$ , 且所在文档为 $\text{doc}_i$
$\text{Event}(\text{tri}_i, \text{label}_i, \text{doc}_i)$	触发词 $\text{tri}_i$ 所在文档为 $\text{doc}_i$ , $\text{label}_i$ 为是否触发了事件
$\text{SameTri}(\text{tri}_i, \text{tri}_j, \text{doc}_i)$	触发词 $\text{tri}_i$ 和 $\text{tri}_j$ 相同, 且它们在同一文档 $\text{doc}_i$ 中
$\text{SameHM}(\text{hm}_i, \text{hm}_j, \text{doc}_i)$	两个触发词所对应的核心词素 $\text{hm}_i$ 和 $\text{hm}_j$ 相同, 且它们在同一文档 $\text{doc}_i$ 中
$\text{SameDoc}(\text{hm}_i, \text{hm}_j, \text{type}_i, \text{doc}_i)$	两个触发词所对应的核心词素分别为 $\text{hm}_i$ 和 $\text{hm}_j$ , 且在同一文档 $\text{doc}_i$ 中, 在 $\text{doc}_i$ 中核心词素为 $\text{hm}_i$ 的候选触发词触发事件的个数与核心词素为 $\text{hm}_i$ 的候选触发词的总数的比值的类型为 $\text{type}_i$

构一样的触发词实例往往有语义上的一致性: 如果该触发词的其中一个实例触发了事件, 那么该文档内其他的含有相同核心词素且触发词形态结构一样的实例有很高概率触发事件。

E4: 武绍祖 26 岁的时候就担任(EM1)全国学联主席, 后来又出任(EM2)国务院副总理王震的秘书, 1985 年更是上任(EM3)国防科工委政委。

**核心词素同指一致性推理(规则(3))** 两个相同形态结构的候选触发词所对应的核心词素  $hm_1$  和  $hm_2$  相同, 且它们在同一文档  $doc_1$  中, 如果核心词素为  $hm_1$  的触发词触发了事件, 那么核心词素为  $hm_2$  的触发词也触发事件。

$$\text{SameHM}(hm_1, hm_2, doc_1) \wedge \text{Event}(hm_1, +label_1, doc_1) \Rightarrow \text{Event}(hm_2, +label_2, doc_1)。(3)$$

## 2.5 相关事件一致性推理

在同一文档中, 相关触发词触发的事件具有一致性。虽然 Die 和 Injure 这两个子类同属于 Life (生活) 大类, Attack 子类属于 Conflict (冲突) 大类, 但这三类事件具有很强的相关性。图 1 显示当 Attack 事件发生时, 其他事件发生的概率。可以看出, Attack 事件与 Injure 事件以及 Die 事件联系非常紧密, 同一文档中, 当 Attack 事件发生时, Die 事件发生的概率是 55%, Injure 事件发生的概率是 43%。例如 E5。

E5: 而在警方和安全部队突击(EM1)伊斯坦堡的温拉尼约监狱的时候, 射(EM2)死(EM3)和第 3 名

朝着军警冲去的自焚囚犯。

E5 中有 3 个事件实例, 包括两个 Attack 事件 (EM1, EM2) 和一个死亡 Die 事件 (EM3)。Attack, Die 和 Injure 这三类事件往往具有相关性: 当其中一类事件发生时, 往往会发生相关事件。

**相关事件一致性推理(规则(4))** 两个触发词所对应的核心词素分别为  $hm_1$  和  $hm_2$ , 且在同一文档  $doc_1$  中, 在  $doc_1$  中核心词素为  $hm_1$  的候选触发词触发事件的个数与核心词素为  $hm_1$  的候选触发词的总数的比值的类型为  $type_1$ , 如果核心词素为  $hm_1$  的候选触发词触发了事件, 那么核心词素为  $hm_2$  的候选触发词也触发事件。

$$\text{SameDoc}(hm_1, hm_2, +type_1, doc_1) \wedge \text{Event}(hm_1, +label_1, doc_1) \Rightarrow \text{Event}(hm_2, +label_2, doc_1)。(4)$$

除 Attack, Die 和 Injure 这三类事件有很强的相关性外, Merge-Org (合并)、Start-Org (创立)、End-Org (倒闭) 和 Declare-Bankruptcy (破产) 这 4 个事件类型也有很强的相关性, 因为它们同属于 Business 事件大类。另外, 属于同一大类的事件类型间也存在相关性 (Movement 事件大类除外), 推理规则 (4) 同样应用于每个事件大类进行推理。

## 2.6 实现方法

本文首先采用 Li 等<sup>[18]</sup>的触发词抽取系统的结果作为推理依据, 采用以下两种方法实现。

1) 对于事件类型一致性推理, 在触发词识别时, 找出在测试集中分类为 0 的测试样本, 形成集合

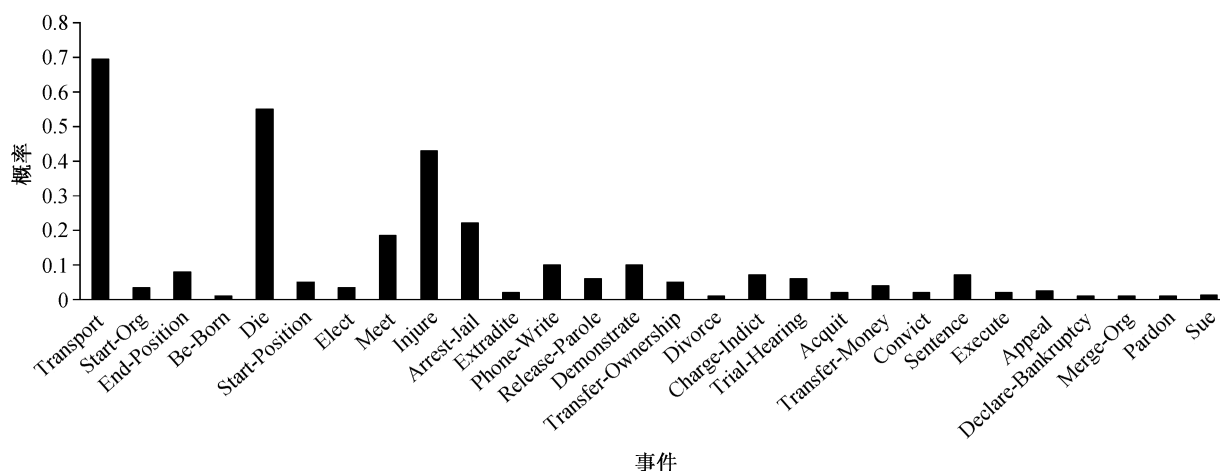


图 1 Attack 事件发生时其他类型事件发生的概率

Fig. 1 Probability of occurrence of other types of events when the Attack event occurs

S1。然后构造 33 个二元分类器对集合 S1 进行分类。如果一个候选触发词在 33 个二元分类器中全部划分为 0, 则该候选触发词为假触发词; 否则选取分类概率最高的类别作为推理时的类别, 形成数据集 S2。在 S2 中, 一个候选触发词在触发词识别时分类的概率为  $P_1$ , 在 33 个二元分类器中最高概率为  $P_2$ , 选择  $P_2$  大于  $P_1$  的候选触发词集合作为推理的测试集。推理训练集为基准系统中的事件类别为 1~33 的触发词。

2) 对于触发词同指一致性、核心词素同指一致性、相关事件一致性推理, 首先提取出候选推理训练测试集。相关事件一致性推理流程如图 2 所示。

首先, 从基准系统训练集中选取目标事件的触发词对应的核心词素, 形成集合  $T$ , 再从训练测试集中选取符合条件的候选触发词。然后, 对这些候选触发词进行过滤, 规则如下: 候选触发词的核心词素必须包含在  $T$  中, 且通过 HowNet<sup>①</sup> 计算候选触发词和对应核心词素的语义相似度, 大于阈值的候选触发词集合最终成为候选推理训练测试集。接着, 在候选推理测试集中设置一个阈值  $V$  (该阈值从开发集中学习得到), 概率大于等于  $V$  的分类结果认为可信, 小于  $V$  的认为不可信。可信部分将作为另外一部分的推理的证据, 不可信部分需要利用推理机制来重新识别, 即作为推理测试集。同时, 还需要建立可信的候选触发词与不可信的候选触发词间的联系。最后, 利用推理训练集和各种证据, 学习一个模型用于推理。触发词同指一致性和核心词素同指一致性两条推理规则的候选推理训练测试集

从基准系统的训练测试集中直接抽取即可得到, 随后的过程与相关事件一致性推理过程相同。

在相关事件一致性推理的 MLN 推理规则中有证据谓词 SameDoc( $hm_1, hm_2, type_1, doc_1$ ), 其中  $type_1$  是这样得来的: 在  $doc_1$  中核心词素为  $hm_1$  的候选触发词触发事件个数与核心词素为  $hm_1$  的候选触发词总数的比值为  $ratio$ , 根据  $ratio$  可以将  $type_1$  划分为  $N$  类, 当  $ratio$  为 0 时(表示核心词素为  $hm_1$  的候选触发词未能触发任何事件),  $type_1$  划分为 0 类, 然后将  $(0, 1]$  的值均分为  $N-1$  类。例如: 选取  $N$  为 3 时, 将  $(0, 1]$  中的值均分为 2 类, 当  $ratio$  为 0 时,  $type_1$  划分为 0 类; 当  $ratio \in (0, 0.5]$  时,  $type_1$  划分为 1 类; 当  $ratio \in (0.5, 1]$  时,  $type_1$  分类为 2 类。如图 3 所示, 在开发集中本文选取的  $N$  从 2 到 7, 当  $N$  为 6 时 F1 值提升最多, 所以本文采用的  $N$  为 6。

### 3 实验

#### 3.1 实验设置

本实验采用 ACE 2005 中文语料, 该语料包含 633 篇中文文档, 包含 8 个事件大类和 33 个事件子类。我们选取 632 篇文档进行实验, 触发词实例总共 3332 个, 随机选取其中的 566 篇作为训练集, 剩余的 66 篇作为测试集, 并且选取训练集中的 33 篇作为开发集。

本实验采用准确率(Precision)、召回率(Recall)及 F1 值作为评测标准, 基准系统使用 Stanford 的最大熵分类器<sup>②</sup>, 参数使用默认值; 推理阶段, 使用

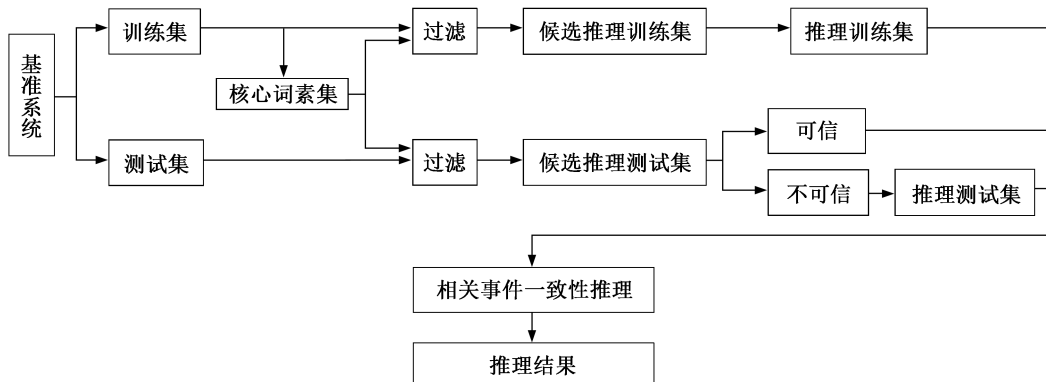


图 2 相关事件一致性推理流程

Fig. 2 Flowchart of inference on relevant events

① <http://www.keenage.com>

② <http://nlp.stanford.edu/software/classifier.shtml>

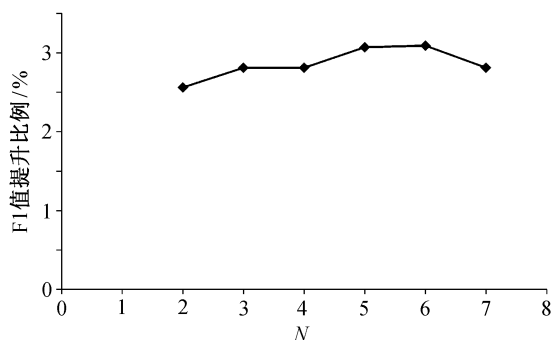


图 3  $N$  值与 F1 值提升的关系  
Fig. 3 The relationship between  $N$  value and improved F1-score

华盛顿大学 MLN 工具 Alchemy<sup>①</sup>; 推理方式使用 MC-SAT。MC-SAT 是一种使用有效的可满足性解决者来进行下一次采样的切片采样马尔科夫链蒙特卡洛算法。

本文的实验以 Li 等<sup>[18]</sup>的自动触发词抽取系统作为基准系统。该基准系统采用 Chen 等<sup>[11]</sup>的特征集, 在触发词识别和事件类型分类时加入以下更加精炼的特征。

1) 句法特征: 候选触发词到句法树根节点的路径。

2) 最近的实体信息: 在候选触发词左边和右边的句法/物理上最近的实体类型和实体。

3) 依存特征: 候选触发词的主语和宾语(当它们为实体时)。

由于存在大量未知的触发词, 在此基础上又提出通过组合语义学和形态结构来扩充触发词。但是

这种扩充方法会引入大量的假触发词, 所以采用触发词过滤机制过滤掉尽可能多的假触发词。本实验使用过滤后的系统作为基准系统。

### 3.2 实验结果及分析

表 2 为基准系统以及加入推理机制后的触发词抽取结果。与基准系统相比, 本文的推理机制对触发词识别和事件类型分类的 F1 值分别提高 3.65% 和 2.51%。

由表 2 可看出, 实验最终提升的总体性能并不是几种推理方法提升性能的简单加和, 这是因为几种推理方法中有相互重叠的部分。比如, 核心词素同指一致性推理和相关事件一致性推理。

经过 MLN 推理后, 触发词识别和事件类型分类的准确率都下降了, 这是因为在推理的时候引入许多噪声, 将本来不是触发词的推理成触发词。召回率都有不小的提升, 说明在分类器阶段有很多真触发词错分为假触发词, 而推理方法将它们重新恢复成真触发词。

加入事件类型一致性推理后, 触发词识别和事件类型分类的 F1 值在基准系统上分别提升 0.92% 和 1.09%, 后者比前者提升多的原因是, 在触发词识别中推理出来的真触发词在事件类型分类中都被正确分类。

加入触发词同指一致性推理后, 触发词识别和事件类型分类的 F1 值都提升不多。这是因为 MLN 推理规则要求两个触发词相同, 并且必须在同一文档中, 这种约束过于严格, 导致符合条件的推理样本过少。

表 2 触发词识别和事件类型分类推理结果  
Table 2 Inference results of trigger identification and event type classification

实验	触发词识别			事件类型分类		
	P/%	R/%	F1/%	P/%	R/%	F1/%
基准系统	72.59	67.85	70.14	65.60	61.31	63.38
+事件类型一致性	(-1.04)	(+2.72)	(+0.92)	(-0.68)	(+2.72)	(+1.09)
+触发词同指一致性	(-0.30)	(+1.09)	(+0.44)	(-0.46)	(+0.82)	(+0.22)
+核心词素同指一致性	(+0.01)	(+2.18)	(+1.15)	(-0.35)	(+1.63)	(+0.69)
+相关事件一致性	(-0.82)	(+4.90)	(+2.12)	(-1.79)	(+3.54)	(+0.95)
本文方法	(70.75) (-1.84)	(77.11) (+9.26)	(73.79) (+3.65)	(63.09) (-2.51)	(68.94) (+7.63)	(65.89) (+2.51)

说明: 括号内数值为提升部分。

① <http://alchemy.cs.washington.edu>

加入核心词素同指一致性推理后, 触发词识别和事件类型分类的 F1 值分别在基准系统上提升 1.15% 和 0.69%。这是因为在触发词同指一致性推理基础上, MLN 推理规则放宽了约束, 只要求两个触发词有相同的核心词素且在同一文档中, 使得符合条件的推理样本数增加很多。例如 E4 中 3 个触发词实例“担任”、“出任”和“升任”的核心词素相同, 可以由触发词实例“担任”触发了事件推理得到“出任”和“升任”也触发了事件。

提升最多的推理规则是相关事件一致性推理规则, 触发词识别和事件类型分类的 F1 值分别提升 2.12% 和 0.95%。这是因为同一文档中事件之间具有相关性, 某一类事件的发生通常伴随着另外一类或几类事件的发生。

相关事件一致性推理中包含 Attack, Die 和 Injure 类相关事件一致性、Business 大类相关事件一致性和非 Movement 大类相关事件一致性推理。

加入 Attack, Die 和 Injure 类相关事件一致性推理后, 在全部类型事件和这三类事件中, 触发词识别性能分别提升 0.97% 和 2.34%, 这是因为这三类事件之间有相关性。例如在同一文档中, 语义为“攻击”的触发词出现时, 语义为“受伤”和“死亡”的触发词出现的概率就很大。而在全部类型事件和这三类事件中, 事件类型分类性能仅分别提升 0.23% 和 0.35%, 这是因为有些触发词语义很丰富。例如触发词“射”可能触发 Attack 类型事件, 也可能触发 Die 或 Injure 类型事件, 造成分类不准确, 从而降低了 F1 值。

加入 Business 大类相关事件一致性推理后, 在全部类型事件中, 触发词识别和事件类型分类的 F1 值分别提升 1.15% 和 0.89%, 在这一大类事件中分别提升 12.08% 和 12.36%。在全部类型事件中提升较少, 在这一大类事件中提升很多, 一方面是因为在测试集中 Business 大类事件的数目较少, 另一方面是因为这一大类下的子类事件之间有相关性。例如在同一文档中, 语义为“设立”的触发词出现时, 语义为“合并”的触发词出现的概率就比较大。

加入非 Movement 大类相关事件一致性推理后, 在全部事件类型中, 触发词识别和事件类型分类分别提升 0.67% 和 0.52%, 在非 Movement 大类事件的所有事件类型下, 分别提升 0.45% 和 0.59%, 提升程度并不明显。这是因为在有些文档中, 事件类型之间的相关性并不明显。例如在同一文档下,

出现 Attack 类型事件的同时也会出现 Start-Org 类型事件, 因此降低了事件类型之间的相关性, 从而影响推理效果。

## 4 总结

本文的主要工作是运用马尔科夫逻辑进行触发词推理, 恢复部分分类阶段未抽取出的触发词实例, 并加入事件类型一致性、触发词同指一致性、核心词素同指一致性和相关事件一致性推理规则。实验结果表明, 系统性能在基准系统的基础上分别提高 3.65% 和 2.51%。

在下一步的研究中, 我们将考虑基于 MLN 的跨文档触发词抽取, 同时也考虑运用 MLN 构建一个联合模型, 进行触发词抽取和论元抽取。

## 参考文献

- [1] Linguistic Data Consortium. ACE (automatic content extraction) Chinese annotation guidelines for events, version 5.5.1 [EB/OL]. (2009-09-08) [2014-03-04]. <http://www.ldc.upenn.edu/Projects/ACE>
- [2] Ahn D. The stages of event extraction // Proceedings of COLING/ACL 2006 Workshop on Annotating and Reasoning about Time and Events. Sydney, 2006: 1-8
- [3] Grishman R, Westbrook D, Meyers A. NYU's English ACE 2005 system description // Proceedings of ACE 2005 Evaluation Workshop. Journal on Satisfiability, 2005, 51(11): 1927-1938
- [4] Hardy H, Kanchakouskaya V, Strzalkowski T. Automatic event classification using surface text features // Proceedings of AAAI 2006 workshop on Event Extraction and Synthesis. Boston, 2006: 36-41
- [5] Finkel J, Grenager T, Manning C. Incorporating non-local information into information extraction systems by Gibbs sampling // Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics. Ann Arbor, 2005: 363-370
- [6] Patwardhan S, Riloff E. A unified model of phrasal and sentential evidence for information extraction // Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing. Singapore, 2009: 151-160
- [7] Ji Heng, Grishman R. Refining event extraction through cross-document inference // Proceedings of

- ACL. Columbus, 2008: 254–262
- [8] Liao Shasha, Grishman R. Using document level cross-event inference to improve event extraction // Proceedings of ACL. Uppsala, 2010: 789–797
- [9] Hong Yu, Zhang Jian, Ma Bin, et al. Using cross-entity inference to improve event extraction // Proceedings of the 49rd Annual Meeting of the Association for Computational Linguistics. Portland, 2011: 1127–1136
- [10] Tan Hongye, Zhao Tiejun, Zheng Jiaheng, et al. Identification of Chinese event and their argument roles // Proceedings of the 2008 IEEE 8th International Conference on Computer and Information Technology Workshops. Sydney, 2008: 14–19
- [11] Chen Zheng, Ji Heng. Language specific issue and feature exploration in Chinese event extraction // Proceedings of 2009 Annual Conference of the North American Chapter of the ACL (NAACL 2009). Boulder, 2009: 209–212
- [12] Fu Jianfeng, Liu Zongtian, Zhong Zhaoman, et al. Chinese event extraction based on feature weighting. Information Technology Journal, 2010, 9(1): 184–187
- [13] Wang Wei. Chinese news event 5W1H semantic elements extraction for event ontology population // Proceedings of WWW 2012. Beijing, 2012: 197–202
- [14] Chen C, Ng V. Joint modeling of Chinese event extraction with rich linguistic features // Proceedings of COLING 2012. Mumbai, 2012: 529–544
- [15] Chen Zheng, Ji Heng. Can one language bootstrap the other: a case study on event extraction // Proceedings of NAACL HLT Workshop on Semi-Supervised Learning for Natural Language Processing. Boulder, 2009: 66–74
- [16] Ji Heng. Cross-lingual predicate cluster acquisition to improve bilingual event extraction by inductive learning // Proceedings of NAACL HLT Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics. Boulder, 2009: 27–35
- [17] Qin Bing, Zhao Yanyan, Ding Xiao, et al. Event type recognition based on trigger expansion. Tsinghua Science and Technology, 2011, 15(3): 251–258
- [18] Li Peifeng, Zhou Guodong, Zhu Qiaoming, et al. Employing compositional semantics and discourse consistency in Chinese event extraction // Proceedings of EMNLP 2012. Jeju Island, 2012: 1006–1016
- [19] Li Peifeng, Zhou Guodong, Zhu Qiaoming, et al. Employing morphological structures and sememes for Chinese event extraction // Proceedings of COLING 2012. Mumbai, 2012: 1619–1634
- [20] Chklovski T, Pantel P. Global path-based refinement of noisy graphs applied to verb semantics // Proceedings of IJCNLP 2005. Berlin, Springer, 2005: 792–803
- [21] Pantel P, Pennacchiotti M. Espresso: leveraging generic patterns for automatically harvesting semantic relations // Proceedings of COLING-ACL 2006. Sydney, 2006: 113–120
- [22] Richardson M, Domingos P. Markov logic networks. Machine Learning, 2006, 62(1/2): 107–136
- [23] Li Peifeng, Zhu Qiaoming, Zhou Guodong. Employing event inference to improve semi-supervised Chinese event extraction // COLING'2014. Dublin, 2014: 2161–2171