

# TimeML在文本时间关系解析中的应用

李路标 张寅生 王惠临  
(中国科学技术信息研究所,北京 100038)

**摘要:** TimeML是在自然语言文本中标注事件和时间表达的一种规范语言。本文介绍了TimeML标准及其语料库TimeBank,从其起源英文文本事件时间表达标注开始,分析TimeML在文本时间关系解析研究中的发展,并探讨了TimeML在汉语文本时间关系解析研究中的应用。

**关键词:** TimeML; TimeBank; 时序推理; 事件识别; 时间信息; 文本时间关系解析

**中图分类号:** G355

**文献标识码:** A

**DOI:** 10.3772/j.issn.1674-1544.2014.05.016

## Application of TimeML in the Text Time Relation Resolution

Li Lubiao, Zhang Yinsheng, Wang Huilin

(Institute of Scientific and Technical Information of China, Beijing 100038)

**Abstract:** TimeML is a robust specification language for event and temporal expressions in natural language text. In recent years, many scholars and researchers have carried on the TRR research from different viewpoints, but not much in Chinese text. In this paper, we give a brief overview of TimeML and TimeBank. Then, from its origin—the annotation of event and temporal expressions in English text, we give a summary and analysis on the previous researches, and provide a new idea for the TRR research in Chinese text.

**Keywords:** TimeML, TimeBank, temporal reasoning, event identification, temporal information, text TRR

### 1 引言

文本时间关系解析(TRR)是自然语言处理领域的一个重要研究内容,主要目的是识别事件及其时间信息的特定关系,并进行事件时间表达关系之间的推理等。它在信息检索系统、问答系统、机器翻译等人工智能领域有着重要的作用。

较好的事件和时间表达式的识别是文本时间关系解析的基础。以往,对事件和时间表达式

的识别主要集中在对词法信息的研究上,而忽略了句子的结构信息和语义信息。随着计算机技术的发展以及大众信息需求的变化,计算机语言学研究者们对文本时间关系中结构信息和语义信息的问题关注增多,而较好的文本事件时间表达标注规范是识别文本事件时间表达关系的基础与进行文本时间关系解析的重要前提。其中最具代表性的是TimeML规范<sup>[1]</sup>。它是在自然语言文本中表示事件和时间表达的一种规范语言<sup>[2]</sup>,是为

---

**作者简介:** 李路标\*(1990-),男,中国科学技术信息研究所硕士研究生,研究方向:自然语言处理;张寅生(1962-),男,中国科学技术信息研究所研究员,博士,主要研究方向:人工智能,科学哲学;王惠临(1948-),男,中国科学技术信息研究所研究员,博士生导师,主要研究方向:多语言信息服务,机器翻译,自然语言处理。

**基金项目:**“十二五”国家科技支撑计划课题“基于多源信息的电动汽车数据挖掘关键技术研究”(2013BAG06B01);国家国际科技合作专项“面向科技文献的日汉双向实用型机器翻译合作研究”(2014DFA11350);中国科学技术信息研究所“自然语言处理”学科建设项目(XK2014-6)。

**收稿日期:**2014年5月12日。

提升自然语言问答系统的处理性能而提出的，现已成为事件时间表达标记的一种ISO标准规范。它被设计用来解决以下4个问题：(1)事件的时间标记；(2)事件之间的相互顺序；(3)根据上下文特定时间表达式对事件进行推理；(4)推理事件的持续时间。TimeML将事件定义为在事件时间关系网中的一个节点，用时间信息来标记和索引事件，从而进行事件在其时间方面的推理等工作<sup>[3]</sup>。这不仅解释了句子的含义，更实现了使计算机理解句子逻辑结构这一语义理解的关键步骤<sup>[4]</sup>。

该标准经过了十几年的发展，基本具有了较为具体和成熟的理论思想。目前，已有不少学者和研究人员从不同的角度基于TimeML标准进行英文文本时间关系解析的研究。在其他外文文中也有针对TimeML的尝试性的实践研究，并有一定进展。但在中文研究领域，并没有深入的展开，相关研究及规范化程度仍然处于初级阶段，无法从句法关系和语义关系的角度对文本中事件及其时间表达关系进行识别分析，也不能实现相关的推理等工作。因此，总结分析前人经验，在前人的基础上将TimeML应用到汉语文本时间关系解析研究的工作上来，服务于中文的问答系统、机器翻译等人工智能领域，将是一项非常有意义的尝试。

## 2 TimeML原理

TimeML是在一个关于问答系统的AQUAINT项目里首次被提出来的。从2002年开始，该标准先后经历了3次专题研讨会的讨论和确定。第一次是TERQAS，研究者们以提高自然语言问答系统的性能，以便处理新闻文章中基于时间

问题的事件为目的，在这次会议上第一次定义TimeML标准并且创建了相应英文文本的语料库TimeBank<sup>[5]</sup>。在接下来的TANGO会议中发布了界面化的标注工具。最后一次会议是TARSQI，在这次会议中参与者们开发了在自然语言文本中表示事件和时间表达的算法，并利用TimeML标准进行了相关的英文文本时间关系解析实验。

TimeML标注形式就是先将待标注文本组织成XML形式，然后用不同的标签来代表不同的元素类别，从而进行文本标注。在TimeML中使用的标签主要有<EVENT><SIGNAL><TIME X3><MAKEINSTANCE>和<LINKS>。其中，<EVENT>标签是用来标注在文本中“事件”类型的元素。在TimeML标注系统中，Boguraev等人认为事件是涵盖“发生”情况的术语<sup>[6]</sup>，既可以是瞬时事件，也可以是持续事件。从语言学角度看，事件可以是动词、名词、形容词等。<TIMEX3>标签是用来标注时间表达式的。<SIGNAL>标签是标注存在于两个实体（事件和事件之间、事件和时间之间、时间和时间之间）之间的具体关系。<MAKEINSTANCE>标签是用来详细描述一个事件的具体实例。<LINKS>标签共有3种类型，分别是<TLINK>、<SLINK>和<ALINK>。<TLINK>（Temporal Link）标签描述的是事件之间、时间之间或者一个事件和时间之间的一种时序关系。<SLINK>（Subordination Link）标签用来描述文本中两个事件之间的关系。<ALINK>（Aspectual Link）标签描述的是一个体态事件和另一个事件之间的联系。

<EVENT>标签和<TIMEX3>标签的BNF表示形式分别如图1和图2，其他标签的BNF表示形式见参考文献[1]。

```

attributes ::= eid class
eid ::= EventID
EventID = e <interger>
class ::= 'OCCURRENCE' | 'PERCEPTION' | 'REPORTING' | 'ASPECTUAL' | 'STATE'
| 'I-STATE' | 'I-ACTION'
comment ::= CDATA
    
```

图1 <EVENT>标签的BNF表示形式

```

attributes ::= tid type [functionInDocument] [beginPoint] [endPoint]
             [quant] [freq] [temporalFunction] (value | valueFromFunction)
             [mod] [anchorTimeID] [comment]
tid ::= ID
{tid ::= TimeID
TimeID ::= t<integer>}
type ::= 'DATE' | 'TIME' | 'DURATION' | 'SET'
beginPoint ::= IDREF
{beginPoint ::= TimeID}
endPoint ::= IDREF
{endPoint ::= TimeID}
quant ::= CDATA
freq ::= Duration
functionInDocument ::= 'CREATION_TIME' | 'EXPIRATION_TIME' | 'MODIFICATION_TIME' |
                       'PUBLICATION_TIME' | 'RELEASE_TIME' | 'RECEPTION_TIME' |
                       'NONE' {default, if absent, is 'NONE'}
temporalFunction ::= 'true' | 'false' {default, if absent, is 'false'}
{temporalFunction ::= boolean}
value ::= Duration | Date | Time | WeekDate | WeekTime | Season | PartOfYear | PaPrFu
valueFromFunction ::= IDREF
{valueFromFunction ::= TemporalFunctionID
TemporalFunctionID ::= tf<integer>}
mod ::= 'BEFORE' | 'AFTER' | 'ON_OR_BEFORE' | 'ON_OR_AFTER' | 'LESS_THAN' | 'MORE_THAN' |
        'EQUAL_OR_LESS' | 'EQUAL_OR_MORE' | 'START' | 'MID' | 'END' | 'APPROX'
anchorTimeID ::= IDREF
{anchorTimeID ::= TimeID}
comment ::= CDATA

```

图2 &lt;TIMEX3&gt;标签的BNF表示形式

在图3中表示的是英文句子“Bill wants to teach on Monday.”的TimeML语言标记结果。

通过TimeML语言的标记,不难发现事件“wants”和“teach”分别标上了<EVENT>标签,时间表达式“Monday”标上了<TIMEX3>标签,时间介词“to”和“on”标上了<SIGNAL>标签,同样还有<TLINK>和<SLINK>等标签。通过这样的标记,不仅可以快速定位到事件及其相关的时间信息,也更便于进行下一步时序推理等研究工作。

随着TimeML在文本时间关系解析研究中的应用,其自身也得到了不断的更新与发展。2004年TimeML1.2版标准发布,不仅对原来1.0版标准进行大幅度的调整,还对相关标签职能分工更加明确,方便进一步处理研究。其后又在此基础上细化了相关属性,扩大了标注范围,发布了TimeML 1.2.1版标准。与之前的事件时间表达式标注规范相比,TimeML 1.2.1版标准具有如下特

征。

(1)在TIMEX2基础上扩大了标注的特征及相关属性值范围。

(2)推出时序功能以便有目的地识别指定的表达式,例如three years ago、last month。

(3)识别决定时间表达的SIGNALS。①时间介词:例如for、during on、at等;②时间连词:例如before、after、while等。

(4)识别所有事件表达的类别。

①时态动词:例如has left、was captured、will resign等;②状态形容词和其他修饰词:例如sunken、stalled、on board等;③事件名词:例如merger、Military Operation、Gulf War等。

(5)创建事件和时间的依存关系。①固定:例如John left on Monday;②排序:例如The party happened after midnight;③嵌入:例如John said Mary left。

可以看出TimeML标准明确了3个在文本事

```

Bill
<EVENT eid="e1" class="I_STATE" >
wants
</EVENT>
<MAKEINSTANCE iid="ei1" eventID="e1" pos="VERB" tense="PRESENT" aspect="NONE"
polarity="POS"/>
<SIGNAL sid="s1">
to
</SIGNAL>
<EVENT eid="e2" class="OCCURRENCE" >
teach
</EVENT>
<MAKEINSTANCE iid="ei2" eventID="e2" pos="VERB" aspect="NONE" tense="INFINITIVE"
polarity="POS"/>
<SIGNAL sid="s2">
on
</SIGNAL>
<TIMEX3 tid="t1" type="DATE" value="XXXX-WXX-1">
Monday
</TIMEX3>
<TLINK eventInstanceID="ei2" signalID="s2" relatedToTime="t1" relType="IS_INCLUDED"/>
<SLINK eventInstanceID="ei1" signalID="s1" subordinatedEventInstance="ei2"
relType="MODAL"/>
    
```

图3 TimeML 语言标记实例

件时间表达标记的不同现象：(1) 系统地将一个事件谓词固定在一个广泛的时间表达式范围内。(2) 对文本中的事件相对于另一个时间表达式进行排序标记。(3) 允许部分表示时间表达式有一个相对延迟的不规范的解释。也正是凭借着较高的表达能力、较大的标注语义范围、较强的模糊问题解释能力以及开源的程序支持，TimeML 标准已被广泛应用到问答系统、机器翻译、信息检索系统等人工智能领域。

### 3 TimeML 在西语研究中的进展

英文的文本时间关系解析研究起步较早。在上世纪90年代，由于缺少成熟的语料库，文本时间关系解析研究大多是基于规则的方法建立起来的事件时间表达关系识别模型，并进行简单地分析工作。1992年，Lascarides、Asher和Oberlander提出了两个事件的时序关系与话语结构有关<sup>[7]</sup>，他们对多种语言环境进行了分析，并对应到narration、elaboration、explanation、background和result

等5种话语关系中。以此，他们根据话语关系和时序关系之间的对应关系，就可以找到事件之间的时序关系。1995年，Hitzeman基于HPSG在分析话语的时序结构时着重考虑了时态(tense)、体态(aspect)、事件副词和修饰关系等因素<sup>[8]</sup>。在2002年，Dorr等人采用基于约束的方法对时态、体态以及连接词进行分析，确定了识别时序关系的约束语言理论模型CONGEN<sup>[9]</sup>。

虽然采用基于规则的方法进行文本时间关系解析研究准确度较高，但是其适用范围小，只能适用较小规模、可形式化的研究内容。随着TimeML标准的明确与完善，尤其是出现了以英文新闻为语料来源的语料库Timebank，同时机器学习的方法也逐渐应用在事件时间表达关系识别上，文本时间关系解析研究有了较为快速的发展，同时也大大推动了相关分析研究工作的进展。

在TimeML标准发布之后，陆续发布了一些用于后续实验研究的基于TimeML标准创建的

语料库, 其中应用最多、具有广泛影响力的当属以英文新闻为语料来源的语料库 TimeBank。语料库 TimeBank 1.1 是在 TimeML 标准诞生的早期、遵循 TimeML 1.1 版标准创建的。较新的语料库 TimeBank 1.2 在 2006 年创建, 它不仅遵循 TimeML 1.2.1 版标准, 而且包含了超过 6.1 万个标记的共计 183 篇相关新闻文章。语料库 TimeBank 代表了最为精细的、具有广泛时序标注的语料库。它不仅是那些对时间和语言感兴趣的语料库语言学家的一份宝贵资源, 也是那些对问答系统、信息抽取等自然语言应用领域感兴趣的语言工程师们的一份宝贵资源。它为今后的研究提供了坚实的基础, 为语义学以及事件时间表达关系推理等研究提供了强有力的支持<sup>[10-11]</sup>。

在 2003 年, Mani 等为了验证机器学习方法的可行性, 他在语料库 Timebank 中抽取时态、时间介词、话语连接词等预设的特征词, 采用监督的方式来训练决策树分类器模型, 不仅达到了 75.4% 的准确率<sup>[12]</sup>, 还进行了初步的新闻事件推理分析研究。2006 年, Mani 等以语料库 Timebank 和语料库 Opinion 为语料来源, 抽取完美特征词, 即 tense、aspect、modality、signal、event class、event string、negation、same tense 和 same aspect, 来训练最大熵分类器模型<sup>[13]</sup>, 并得到了 62.5% 的准确率。他也提出了用时序推理的方法来扩大训练语料中 <TLINK> 等标签的数量来解决语料库数据稀疏的问题。在 2007 年, Chambers 等在 Mani 的研究基础上, 扩大了特征空间, 加入了词性等特征, 提升了分类器的准确率, 使之达到了 67.57%<sup>[14]</sup>。在此之中, 他发现两个事件之间的依赖关系特征对分类器有着重要的影响, 并提出了一种基于纯文本的事件时间表达关系识别的方法, 分为两个阶段——自动标注事件属性阶段和事件时间表达关系识别阶段, 这大大解决了在事件时间表达关系识别研究中的数据稀疏问题。

在 2008 年, Chambers 等又提出了解决分类器测试结果后的事件时间表达关系冲突问题来提高文本时间关系分析模型性能的观点<sup>[15]</sup>。这之前通过提高分类器的准确率来提高文本时间关系

分析模型性能的普遍做法做了调整, 他们用 ILP (Integer Linear Programming) 来限制结果, 如果发现冲突则用推理的方法重新进行事件时间表达关系识别与分析, 但是这一方法是比较繁琐的。Yoshikawa 在 2009 年提出了用马尔科夫模型来进行事件时间表达关系识别的观点<sup>[16]</sup>, 并有效避免了用 ILP 模型检测的繁琐。

针对 TimeML 中 <TIMEX3> 标签的升级, 在 2011 年, Saquete 和 Pustejovsky 实现了在 TIDES 中从 <TIMEX2> 标签到 <TIMEX3> 的自动转化, 并达到了将近 90% 的准确率<sup>[17]</sup>。2012 年, Derczynski Leon 和 Gaizauskas Robert 利用 TimeML 中 <SIGNAL> 标签对时序关系分类问题进行了研究, 并提升了其准确率<sup>[18]</sup>。Chang A X 和 Manning C 按照 TimeML 标准, 开发了英文文本中时间表达式的识别和规范化处理的 SUTime<sup>[19]</sup>, SUTime 有着较高的识别准确率, 现在已经纳入斯坦福大学自然语言处理项目。就在 2013 年刚刚结束的关于词法和计算语义学第二次联合会议上, 众多自然语言处理领域的学者和研究人员聚集在一起, 深入研讨了基于 TimeML 的文本时间关系解析研究情况。这次不仅进行了新一轮的评估工作, 更重要的是实现了更进一步的解析研究。Chambers N 实现了从生语料文件中对事件和时间表达式的排序处理, NavyTime 在对时间表达式进行识别与分析时超越了 SUTime, 使得文本时间关系解析研究更加向前推进了一步<sup>[20]</sup>。Wartena Christian 利用支持向量机对词和短语进行了语义相似度的评估<sup>[21]</sup>。Lau Jey Han、Cook Paul 和 Baldwin Timothy 基于主题模型进行了词义归纳研究<sup>[22]</sup>。

国内也有一些基于 TimeML 标准进行英文文本时间关系解析的研究。2010 年, 孙辉结合机器学习的方法和计算语言学的知识构造了基于 OTC 语料库的英文事件时序关系识别模型, 不仅解决了语料库数据稀疏问题, 也尝试性地应用在其他大规模的语料库中<sup>[23]</sup>。在 2012 年, 王凤玲提出了使用 CRF 模型结合多种特征方法对英语时间表达式进行识别与分析的观点<sup>[24]</sup>, 并采用 TimeBank1.1 为评测语料进行实验, 不仅验证了

其观点,也得到了一些理想的效果。

基于TimeML进行英文文本时间关系解析的研究已经较为成熟,从最初仅仅依靠规则进行事件、时间表达等单方面的识别分析,到结合语料库TimeBank、利用机器学习的方法进行时序推理等全方面、深层次的研究,英文文本时间关系解析研究已遥遥领先。针对具体的研究任务,其研究方法也不尽相同,即便是相同或者相似的研究任务,也会有不少新意。总体来说,针对英文事件、时间表达等识别与分析的研究方法已基本成型,即通过构建分类器模型,训练已标注好的文本,进而对生语料进行识别与分析。但是,在进一步的时序推理研究中并没有一致的方法:一是针对不同的研究内容,具体的方法不同;二是在TimeML标准发展的短短10来年的时间里,研究工作者的研究内容不是十分集中,比较分散,相应的研究方法也不统一。

随着应用TimeML标准进行时间关系解析研究热度的上升,越来越多的学者和研究人员开始探索将TimeML标准应用在其他语言上。2007年, Frank在英德平行语料基础上,应用词对齐技术实现了德语的TimeML时间表达标注<sup>[25]</sup>。他利用现有的标注工具及平台,标注英德平行语料中的英语部分,然后利用词对齐技术,自动将这些标注映射到德语文本。在2009年, Caselli T等成功地开发出基于TimeML的规范化的意大利语言文本事件时间表达关系的标注规范<sup>[26]</sup>。他们采用的是在时间表达之间基于WordNet的语义关系模型,并达到了86.41%的准确率。在2010年, Saur Roser给出了基于TimeML标准涵盖加泰罗尼亚语和西班牙语时序关系的标注规范<sup>[27]</sup>。2011年, Bittar André等在语言现象等问题上对TimeML标准进行了修改和提升,构建了法语的语料库(French TimeBank)<sup>[28]</sup>。同样还有将TimeML标准应用在其他语言来进行文本时间关系解析的研究<sup>[29-31]</sup>,这些无不说明了TimeML标准强大的适用性及应用前景。

近几年,基于TimeML进行其他外文文本时间关系解析的研究热度呈上升趋势。由于不同语

种的语言结构不同,使得TimeML并不能完全直接应用在其语言上。但是,通过其他外文的成功应用可以看出,借助词对齐的映射方法是一个不错的选择,一方面词对齐技术作为一较为成熟的方法已经被广泛应用在双语语料的实践中,另一方面可以充分利用现有的平台及相应的英文资源。无论在实现基于TimeML进行目标语言的标注上,还是在构建目标语言相应的语料库上,借助词对齐的映射模型方法都起到了重要作用。

#### 4 TimeML在中文研究中的进展

中文的文本时间关系解析研究起步较晚,由于中文与英文语法结构的差异性及汉语现象的繁多性,甚至到现在也没有基于TimeML的中文语料库。

Li Wenjie和Wong Kam-Fai是中文文本时间关系解析研究的先驱者。他们在2002年提出了基于规则的事件时间表达关系识别与分析的方法<sup>[32]</sup>。他们以同一个句子中的两个事件为研究对象,仅仅考虑像before、after这样的连接词,设计了一系列从时序关系连接到时序关系的映射方案。这种方法虽然简单,但效率不高,难以推广。鉴于上述方法的各种缺陷,他们在2004年提出了用机器学习的方法来进行汉语文本时间关系解析的研究<sup>[33]</sup>。他们从香港中文报纸的金融板块上抽取了700个句子,手工标注其中的600个句子,在标注的600个句子中,用400个句子来进行模型的训练,100个句子用来测试,100个句子留存。他们采用了很多像时态(tense)、体态(aspect)等与英文事件时间表达关系识别相似的特征,并用贝叶斯分类器进行实验,达到了82.25%的准确率。但是他们的语料构建和特征的选取全部是通过手工进行的,也没有进行事件与时间表达式之间的相关分析。

2007年, Cheng Yuchang等在TimeML标准基础上提出了用词语间的依赖关系来进行事件时间表达关系识别的方法,并建立了用于中文的事件时间表达关系识别的“语料库”<sup>[34]</sup>。为了减少人工标注的工作量,他优化了一些规范,不仅限制了

事件只能是动词,而且也限制了时间联系的相关事件。他认为只有符合“相邻的事件”“语法树中具有父子关系的事件”和“兄弟节点的事件”这三种依赖关系之一,才认为事件可能发生时间上的联系。依照这种标注规则,2008年他采用统计机器学习的方法对从Penn Chinese TreeBank中抽取的10%数据构建的语料库进行实验与分析,并提出了包括RLP (Relation to Linear Preceding event)、RTA (Relation to Tree Ancestor event)和RTP (Relation to Tree Preceding event)的事件类型这一属性来完善语料库信息<sup>[35]</sup>。实验结果表明,加入的事件类型信息提高了分类器支持向量机(Support Vector Machine, SVM)的准确率,但遗憾的是实验中并没有涉及语义等属性。

TimeML 1.2.1 版标准的发布,给从事文本时间关系解析研究的学者和研究人员带来了新的生机。在2011年,Llorens H等基于语义角色来进行中文事件时间表达关系识别的初步研究,并证明了此方法的可行性<sup>[36]</sup>。2012年,Entrikin R在对汉语文本中体态问题进行研究时,探索了基于语料库语言学技术来进行汉语文本中结构语法的检查,并且采用基于语料库的机器学习方法分别训练了汉语文本体态标注用法的最大熵分类器和条件随机域模型,成功地将其应用在汉语文本体态标注中<sup>[37]</sup>。但是,由于过度集中在对模型的训练上,而忽视了对构建模型的评价以及标注策略的优化。在2013年,Zhang Xujie等针对TimeML中定义的事件类型,利用支持向量机对中文的事件进行分类研究,达到了81.16%的准确率。但是他们在构建事件抽取规则时只选取了POS和位置信息(position information)等属性,并没有涉及更多的语义属性<sup>[38]</sup>。

基于TimeML进行中文文本时间关系解析研究才刚刚起步,相关研究以及平台的支持还在进一步的探索中。从现有的工作进展中,可以看到,在对汉语事件、时间表达等识别与分析中还存在一些不足,比如:严格限制汉语事件类型的识别;不能实现汉语事件时间表达关系的自动标注,也不能很好地进行深层次的时序推理研

究等。但是,前人的研究工作也给了继续探索TimeML在汉语上应用的很好的指示:一是确定了基于TimeML标准进行汉语文本时间关系解析研究的可行性,二是坚定了后来人继续探索、继续前进的信心。

## 5 总结与展望

(1)从TimeML在文本时间关系解析研究中的发展来看,各个工作之间都有着紧密的联系和延续性。TimeML 1.1版标准提供了标注文本中事件时间表达信息的框架,为事件时间表达识别研究提供了一种新的方法。TimeML 1.2及1.2.1版本标准细化了使用的标签及其相关属性,扩大了标注范围,提高了文本标注精度,诠释了TimeML更加强大的标注能力,为文本时间关系解析研究提供了更好的标注规范及相关平台支持。

(2)语料库TimeBank 1.1和TimeBank 1.2不管是在基于规则方法进行文本时间关系解析研究的初始阶段,还是在基于机器学习方法进行文本时间关系解析研究的发展阶段,都给予了强有力的资源支持。

(3)利用现有英文研究的相关平台及资源,借鉴德语、法语研究的映射方法及技术,可以找到TimeML更加契合中文文本时间关系解析的研究方法,使其服务于中文的问答系统、机器翻译等人工智能领域。

(4)具体说来,结合语料库TimeBank,借助机器学习的方法训练合适的分类器引擎,实现对英汉平行语料中英文部分的自动标注;利用词对齐技术,构建英汉映射模型,将其映射到汉语部分,实现汉语文本事件时间表达关系的识别与分析,并构建汉语文本事件时间表达关系语料库;基于汉语单语料构建汉语文本时间关系解析引擎,从而进行汉语时序分析等研究。

(5)在今后的研究工作中,可以尝试通过英汉平行语料,基于词对齐的映射方法和技术,构建符合TimeML标准的事件时间表达关系的汉语语料库,并结合机器学习的方法,训练汉语文本

时间关系解析引擎, 从而进行汉语文本时间关系解析实验研究。

### 参考文献

- [1] Saurii R, Littman J, Knippen B, et al. TimeML Annotation Guidelines[J/OL]. [2014-03-22]. <http://www.cs.brandeis.edu/~cs112/cs112-2004/annPS/annguide-12wptoc.pdf>.
- [2] Pustejovsky J, Castano J M, Ingria R, et al. TimeML: Robust Specification of Event and Temporal Expressions in Text[J]. *New Directions in Question Answering*, 2003(3): 28-34.
- [3] Boguraev B, Ando R K. TimeML-Compliant Text Analysis for Temporal Reasoning[C]. *IJCAI*, 2005, 5: 997-1003.
- [4] Kumari G V, Sanampudi S K. Temporal Reasoning in Natural Language Processing: A Survey[J]. *International Journal of Computer Applications*, 2010(4): 68-72.
- [5] Pustejovsky J, Hanks P, Sauri R, et al. The Timebank Corpus[C]. *Corpus Linguistics*, 2003: 40.
- [6] Boguraev Branimir, Castaño Jose, Gaizauskas, et al. TimeML 1.2.1 Specifications[J/OL]. [2014-03-22]. [http://timeml.org/site/publications/timeMLdocs/timeml\\_1.2.1.html](http://timeml.org/site/publications/timeMLdocs/timeml_1.2.1.html).
- [7] Lascarides A, Asher N, Oberlander J. Inferring Discourse Relations in Context[C]//*Proceedings of the 30th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1992: 1-8.
- [8] Hitzeman J, Moens M, Grover C. Algorithms for Analyzing the Temporal Structure of Discourse[C]. *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics*. Burlington: Morgan Kaufmann Publishers Inc., 1995: 253-260.
- [9] Dorr B J, Gaasterland T. Constraints on the Generation of Tense, Aspect, and Connecting Words from Temporal Expressions[J]. *Journal of Artificial Intelligence Research (JAIR)*, 2002.
- [10] Boguraev B, Ando R K. TimeBank Driven TimeML-Analysis[J]. *Annotating, Extracting and Reasoning about Time and Events*, 2005.
- [11] Boguraev B, Pustejovsky J, Ando R, et al. Timebank Evolution as a Community Resource for Timeml Parsing[J]. *Language Resources and Evaluation*, 2007, 41(1): 91-115.
- [12] Mani I, Schiffman B, Zhang J. Inferring Temporal Ordering of Events in News[C]//*Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Companion Volume of the Proceedings of HLT-NAACL 2003--Short Papers--Volume 2*. Association for Computational Linguistics, 2003: 55-57.
- [13] Mani I, Verhagen M, Wellner B, et al. Machine Learning of Temporal Relations[C]//*Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006: 753-760.
- [14] Chambers N, Wang S, Jurafsky D. Classifying Temporal Relations between Events[C]//*Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*. Association for Computational Linguistics, 2007: 173-176.
- [15] Chambers N, Jurafsky D. Jointly Combining Implicit Constraints Improves Temporal Ordering[C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2008: 698-706.
- [16] Yoshikawa K, Riedel S, Asahara M, et al. Jointly Identifying Temporal Relations with Markov Logic[C]//*Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1--Volume 1*. Association for Computational Linguistics, 2009: 405-413.
- [17] Saquete E, Pustejovsky J. Automatic Transformation from TIDES to TimeMLannotation[J]. *Language Resources and Evaluation*, 2011, 45(4): 495-523.
- [18] Derczynski L, Gaizauskas R. Using Signals to Improve Automatic Classification of Temporal Relations[J/OL]. *arXiv e-print*, 2012[2014-03-22].<http://arXiv.org/pdf/1203.50551.pdf>.
- [19] Chang A X, Manning C. SUTime: A Library for Recognizing and Normalizing Time Expressions[C]. *LREC*. 2012: 3735-3740.
- [20] Chambers N. Navytime: Event and Time Ordering from Raw Text[C]. *Second Joint Conference on Lexical and Computational Semantics (\* SEM)*, 2013, 2: 73-77.

- [21] Wartena C. HsH: Estimating Semantic Similarity of Words and Short Phrases with Frequency Normalized Distance Measures[C]// Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013).Atlanta, Georgia, USA, 2013: 48.
- [22] Lau J H, Cook P, Baldwin T. Unimelb: Topic Modelling-based Word Sense Induction[C]. Second Joint Conference on Lexical and Computational Semantics (\*SEM), 2013, 2: 307-311.
- [23] 孙辉.事件时序关系识别的研究与实现[D].哈尔滨:哈尔滨工业大学, 2010.
- [24] 王凤玲.基于条件随机域模型的英语时间表达式识别研究[J].电子技术,2012, 39(5): 8-10.
- [25] Frank P D D A. Projecting Temporal Annotations Across Languages[D]. Saarbrücken, Germany: Universität des Saarlandes, 2007.
- [26] Caselli T, Prodanof I. TETI: A TimeML Compliant TimeEx Tagger for Italian[C]// Computer Science and Information Technology, 2009. IMCSIT'09. International Multiconference on. IEEE, 2009: 185-192.
- [27] Sauri R. Annotating Temporal Relations in Catalan and Spanish.TimeML Annotation Guidelines (Version TempEval-2010)[J/OL]. [2014-03-22]. [http://comunicacio.barcelonamedia.org/technical\\_reports/BM2010\\_04.pdf](http://comunicacio.barcelonamedia.org/technical_reports/BM2010_04.pdf).
- [28] Bittar A, Amsili P, Denis P, et al. French TimeBank: An ISO-TimeML Annotated Reference Corpus[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (Short Papers), 2011: 130-134.
- [29] Spreyer K, Frank A. Projection-based Acquisition of a Temporal Labeller[C]. IJCNLP, 2008: 489-496.
- [30] Sauri R, Badia T. Spanish TimeBank 1.0 Corpus documentation[J/OL]. [2014-03-22]. [https://catalog.ldc.upenn.edu/docs/LDC2012T12/README\\_spanish-TimeBank.pdf](https://catalog.ldc.upenn.edu/docs/LDC2012T12/README_spanish-TimeBank.pdf).
- [31] Jarzębowski P, Przepiórkowski A. Temporal Information Extraction with Cross-Language Projected Data[M]// Advances in Natural Language Processing. Berlin, Germany: Springer Berlin Heidelberg, 2012: 198-209.
- [32] Li W, Wong K F. A Word-based Approach for Modeling and Discovering Temporal Relations Embedded in Chinese Sentences[J]. ACM Transactions on Asian Language Information Processing (TALIP), 2002(3): 173-206.
- [33] Li W, Wong K F, Cao G, et al. Applying Machine Learning to Chinese Temporal Relation Resolution[C]// Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004: 582.
- [34] Cheng Y, Asahara M, Matsumoto Y. Constructing a Temporal Relation Tagged Corpus of Chinese Based on Dependency Structure Analysis[C]//Temporal Representation and Reasoning, 14th International Symposium on. IEEE, 2007: 59-69.
- [35] Cheng Y, Asahara M, Matsumoto Y. Use of Event Types for Temporal Relation Identification in Chinese Text[C]. IJCNLP, 2008: 31-38.
- [36] Llorens H, Saquete E, Navarro B, et al. Data-driven Approach Based on Semantic Roles for Recognizing Temporal Expressions and Events in Chinese[M]// Natural Language Processing and Information Systems. Berlin, Germany: Springer Berlin Heidelberg, 2011: 88-99.
- [37] Entrikin R. Applying Machine Learning to Usage of Aspect Markers in Chinese Text[D]. Massachusetts: Brandeis University, 2012.
- [38] Zhang X, Liu Z, Liu W, et al. Chinese Event Classification for Event Ontology Construction[J]. Journal of Computational Information Systems, 2013, 9(9): 3511-3519.

(上接第83页)

- [13] Qin J, Li K. How Portable Are the Metadata Standards for Scientific Data? A Proposal for a Metadata Infrastructure[EB/OL]. [2013-10-28]. <http://dcevents.dublincore.org/IntConf/index/pages/view/2013-peer-Abstracts#Qin>.
- [14] 孟宪学,赵瑞雪,周国民,等.农业科学数据核心元数据[M].北京:中国农业科学技术出版社, 2005.
- [15] You S. Evaluative Metadata in Educational Digital Libraries: How Users Use Evaluative Metadata in the Process of Document Selection[J]. TCDL Bulletin, 2010, 4(2): 1-11.