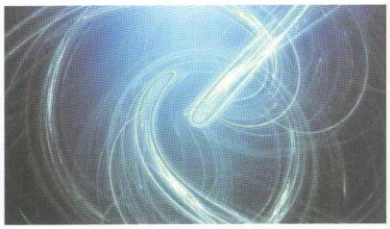
国家期刊奖提名奖
中国高校精品科技期刊奖
全国高校科技期刊优秀编辑质量奖
百种中国杰出学术期刊奖
湖北省优秀期刊奖

ISSN 1671-8836
CODEN WDXLA5

武漢大學學報（理学版）

Journal of Wuhan University (Natural Science Edition)



第61卷 第2期

ISSN 1671-8836

2015

武汉大学学报(理学版)
2015年第61卷第2期(总第270期)

目　次

责任编辑：谭　辉
编　　辑：曹启花　田志东　胡　敏　卢佳华
出版日期：2015年4月24日

[期刊基本参数] CN42-1674/N * 1930 * b * A4 * 96 * zh * P * ¥20.00 * 1150 * 16 * 2015-04

第 61 卷 第 2 期
2015 年 4 月

武汉大学学报(理学版)
J. Wuhan Univ. (Nat. Sci. Ed.)

Vol. 61 No. 2
Apr. 2015,139~144

DOI:10.14188/j.1671-8836.2015.02.006

基于信息单元融合的新闻原子事件抽取

张 贺[1,2],刘茂福[1,2†],胡慧君[1,2],顾进广[1,2]

(1. 武汉科技大学 计算机科学与技术学院,湖北 武汉 430065;
2. 智能信息处理与实时工业系统湖北省重点实验室,湖北 武汉 430065)

摘 要:原子事件抽取是将非结构化文本进行结构化表示的重要方法.针对新闻语料,本文提出了一种基于信息单元融合的原子事件抽取方法.在中文分词、词性标注、命名实体识别等自然语言处理技术的基础上,利用语言规则将信息单元标识出来并进行融合,达到浅层句法分析的效果,通过原子事件抽取算法将原子事件从经信息单元融合后的语料中抽取出来.基于信息单元融合的原子事件抽取方法不仅对文本长度没有严格限制,并且不受事件类型的约束;实验结果表明,基于信息单元融合的原子事件抽取方法是有效的.

关 键 词:信息单元融合;原子事件;事件抽取

中图分类号:TP 391 **文献标识码**:A **文章编号**:1671-8836(2015)02-0139-06

Atomic Event Extraction Based on Information Unit Fusion

ZHANG He[1,2], LIU Maofu[1,2†], HU Huijun[1,2], GU Jinguang[1,2]

(1. College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China; 2. Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan 430065, Hubei, China)

Abstract: Atomic event extraction is an important means to represent the unstructured text structurally. This paper proposes an information basic unit fusion approach to extract atomic event from the news. On the basis of Chinese word segmentation, part of speech tagging and named entity recognition, information basic units can be marked and fused according to linguistic rules. And then, the atomic events can be extracted from information basic unit fused texts by the atomic event extraction algorithm. This approach does not restrict the length of texts and the types of atomic events. The experiment results demonstrate the effectiveness and feasibility of the atomic event extraction approach based on information unit fusion.

Key words: information unit fusion; atomic event; event extraction

0 引 言

事件抽取隶属于信息抽取领域,主要研究如何从含有事件信息的非结构化文本中提取出用户感兴趣的事件信息,把以自然语言形式表达的事件以结构化的形式呈现出来,如什么人、在什么地方、什么时间做了什么事等[1].事件抽取的研究是多学科发展和应用的需要,在自动文摘[2],信息检索[3]等领域均有广泛应用.

事件抽取主要包括原子事件抽取与主题事件抽取两类.目前大多数关于事件抽取方面的研究都是主题事件研究,本文提出的方法属于原子事件抽取范畴.事件抽取方法主要有基于模式匹配的方法和基于机器学习的方法.基于模式匹配的方法,文献[4]提出基于谓词论元结构的中文事件抽取方法,根据文本的语法结构和谓词论元结构,找出动词支配的论元,最后利用语义属性确定相应事件特征;侯立斌等提出基于跨事件的缺失事件角色填充理论[5],参考 Gupta 与 Ji[6] 制定的规则,基于缺失角色的分布情况,设定了一系列规则来对事件缺失角色进行填充.基于机器学习的方法,Llorens 等通过 CRF 模

收稿日期:2014-08-30 †通信联系人 E-mail:liumaofu@wust.edu.cn
基金项目:国家自然科学基金(61100133,61173062),国家社会科学基金重大项目(11&ZD189)
作者简介:张 贺,男,硕士生,主要从事自然语言处理方面的研究. E-mail:cheesezh@qq.com

型进行语义角色标注[7]，并应用于 TimeML 的事件抽取，提升了系统性能；Ahn 结合 MegaM 和 TiMBL 两种机器学习方法分别实现了事件类别识别和事件元素识别两大任务[8]；赵妍妍等采用一种基于触发词扩展和二元分类相结合的方法来识别事件类别[1]，以此解决了触发词导致的训练中正反例不平衡问题；徐红磊和许旭阳等采用基于事件实例的方式进行事件的探测[9-10]，这种方法克服了传统基于触发词方法中正反例失衡以及数据稀疏问题.

基于模式匹配的方法在特定领域内可以取得比较好的效果，但是系统的可移植性差. 基于机器学习的方法虽然不依赖于语料内容与格式，但是需要大规模的标注语料，否则会出现较为严重的数据稀疏问题. 本文采用基于信息单元融合的方法进行事件抽取，实质是一种基于规则的方法，抽取对象是新闻语料中符合事件结构的原子事件而并不局限于某一类型或某一领域的特定事件，具有较好的普适性，不需要用已标注语料进行训练，简单实用. 此外，事件抽取的理想情况是在深层语义分析的基础上进行，但是目前深层语义分析技术尚不成熟，不仅准确率偏低以至于并不实用，而且运行效率低下，对文本长度限制较为严格. 基于信息单元融合的方法对文本的分析介于浅层语法分析和深层语义分析之间，无需对文本内部细节进行精细的分析，而且对文本长度限制也比深层句法分析宽松许多，是一种更加实用、高效、便捷的方法.

文献[11]将事件抽取的事件结构中的语义角色分为核心语义角色和附加语义角色两类. 核心语义角色又包括主体、客体和时空三类，附加语义角色有工具、方式、材料、原因和范围五类原子语义角色. 为了简化事件的基本结构，本文将事件结构定义为谓词、主体、客体、时间、地点五元组.

1 系统结构

在基于信息单元融合的新闻原子事件抽取方法中，经过源文档预处理、动词过滤、初步融合、信息单元融合、核心度计算、原子事件抽取等六个阶段的处理后，可以抽取出源文档中符合事件结构的原子事件. 具体的系统结构如图 1 所示.

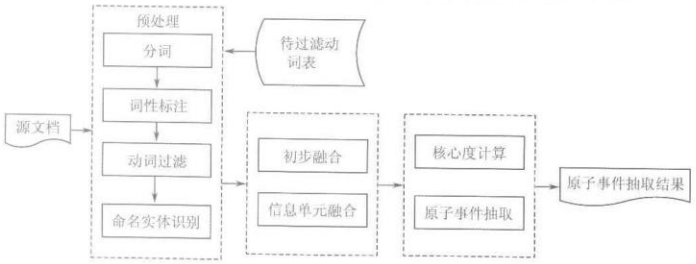


图 1 基于信息单元融合的新闻原子事件抽取系统结构图

在源文档预处理阶段，首先利用斯坦福自然语言处理工具(*http://nlp.stanford.edu/software/index.shtml*)对源文档进行分词，然后在分词结果的基础上分别进行词性标注和命名实体识别. 由于原子事件的抽取是由谓词驱动，并且本文认为当一个分句中存在嵌套事件或者并列事件时只有一个核心事件，因此为了抽取出核心事件，需要过滤掉分句中非核心事件的谓词. 通过对大量语料进行分析，我们发现意向动词、言说动词、趋向动词以及致使动词这四类动词在使用的过程中，当与其他动词存在于同一上下文时，动词的词性会被虚化. 我们在待过滤动词表中总结并整理了这四类动词，当这四类动词与其他动词共现于同一分句时，需要被过滤掉.

初步融合、信息单元融合、核心度计算和原子事件抽取会在后面的章节详细叙述.

2 原子事件抽取方法

2.1 初步融合规则库

新闻语料中经常会出现人名、地名等名词性信息单元，这类信息单元在中分文词过程中被分解并且在词性标注过程中被标记为名词、代词，这对事件抽取的准确性有一定的影响并且使计算机不能清晰的理解语料的含义，而命名实体识别的功能是识别出一段语料中的人物、地点和地缘政治实体等命名实体，这恰好弥补了词性标注的不足，因此需要用命名实体在命名实体识别结果中的标签替换掉该词在词性标注结果中的词性标签. 此外，由于用书名号所

标识的成分在分词过程中被分解,在初步融合的过程中需要把书名号所标识的成分进行合并并且标记为名词.表1中示例1至2为初步融合结果示例.文中词性标签和命名实体识别标签含义与 Penn Chinese Treebank Tag Set[13]中的标签含义相同,本文出现的标签及解释如表2所示.

表1　初步融合结果示例表

示例1	词性标注结果	李宗盛/NR,在/P,北京/NR,举办/VV,演唱会/NN,,/PU
	命名实体识别结果	李宗盛/PERSON 在/O 北京/GPE 举办/O 演唱会/O ./O
	初步融合结果	李宗盛/ PERSON,在/P,北京/ GPE,举办/VV,演唱会/NN,,/PU
示例2	词性标注结果	《/PU 统计/NN 自然/NN 语言/NN 处理/NN 》/PU
	初步融合结果	《统计自然语言处理》/NN

表2　本文中标签及解释

标签	解释	标签	解释
CC	连词	CD	数词
DEG	关联词"的"	DT	代词
LC	方位词	M	量词
NN	普通名词	NR	专有名词
P	介词	PN	人称代词
PU	标点符号	VV	动词
JJ	名词修饰词	PERSON	人物实体
LOC	地点实体	GPE	地缘政治实体
O	非命名实体		

2.2　信息单元融合规则库

信息单元融合规则库的构建是本方法的关键.初步融合后的文本对于事件抽取而言存在切分粒度过细的问题,因此从中准确识别事件成分是比较困难的.此外,初步融合后的文本中所含标签也是词性标签和命名实体标签,因此在一个语句中会存在多个动词和名词,这也给事件抽取带来较大困难.信息单元融合的方法可以有效解决这两个问题.通过对大量语料进行分析,我们将发现的语言知识归纳到信息单元融合规则库中.信息单元融合规则库主要包括修饰语单元融合规则、人名实体单元融合规则、名词单元融合规则和状语单元融合规则四类.根据修饰语单元融合规则、人名实体单元融合规则、名词单元融合规则和状语单元融合规则,将初步融合后的文本进行信息单元融合,信息单元融合后的文本就包含了介于浅层语法分析和深层语义分析之间的信息单元标签,一个信息单元就是一个独立完整的事件成分,在后续的原子事件抽取过程中可以直接利用信息单元标签的信息进行事件成分的抽取.

为了便于利用规则进行信息单元融合,本文将初步融合后的文本用词和标签组成的序列 T 表示:$T=(w_1/t_1, w_2/t_2, w_3/t_3, \cdots, w_n/t_n)$,其中 w_i 表示文本序列中的一个词,t_i 表示 w_i 对应的词性标签或命名实体标签.

1) 修饰语单元融合规则

设标签集合 $P=\{$NN, NR, PN, VV, JJ, PERSON, GPE$\}$,标签集合 $Q=\{$DEG, JJ$\}$,标签集合 $R=\{$JJ$\}$.当文本中相邻的两个元素 w_i/t_i 与 w_{i+1}/t_{i+1} 的标签之间满足"$t_i \in P$ 且 $t_{i+1} \in Q$",则将 w_i 与 w_{i+1} 进行融合并赋予修饰语单元标签 MIU;当文本中相邻的三个元素 w_{i-1}/t_{i-1}、w_i/t_i 以及 w_{i+1}/t_{i+1} 的标签之间满足"$t_{i-1} \in P$ 且 $t_i \in Q$ 且 $t_{i+1} \in R$",则将 w_{i-1}、w_i 以及 w_{i+1} 进行融合并赋予修饰语单元标签 MIU.修饰语单元融合示例如表3中示例3至5所示.

2) 人名实体单元融合规则

设标签集合 $P=\{$NN, NR, PN, JJ, PERSON, MIU$\}$,标签集合 $Q=\{$CC, PN, PERSON$\}$,标签集合 $R=\{$PN, PERSON$\}$.当文本中相邻的两个元素 w_i/t_i 与 w_{i+1}/t_{i+1} 的标签之间满足"$t_i \in P$ 且 $t_{i+1} \in Q$",则将 w_i 与 w_{i+1} 进行融合并赋予人名实体单元标签 PEIU;当文本中相邻的三个元素 w_{i-1}/t_{i-1}、w_i/t_i 以及 w_{i+1}/t_{i+1} 的标签之间满足"$t_{i-1} \in P$ 且 $t_i \in Q$ 且 $t_{i+1} \in R$",则将 w_{i-1}、w_i 以及 w_{i+1} 进行融合并赋予人名实体单元标签 PEIU.人名实体单元融合示例如表3中示例6所示.

3) 名词单元融合规则

设标签集合 $P=\{$CD, DT, NN, NR, JJ, PN, PERSON, LOC, GPE, MIU$\}$,标签集合 $Q=\{$CC, CD, DEG, M, NN, NR, MIU$\}$,标签集合 $R=\{$MIU, NN, NR$\}$.当文本中相邻的两个元素 w_i/t_i 与 w_{i+1}/t_{i+1} 的标签之间满足"$t_i \in P$ 且 $t_{i+1} \in Q$",则将 w_i 与 w_{i+1} 进行融合并赋予名词单元单元标签 NIU;当文本中相邻的三个元素 w_{i-1}/t_{i-1}、w_i/t_i 以及 w_{i+1}/t_{i+1} 的标签之间满足"$t_{i-1} \in P$ 且 $t_i \in Q$ 且 $t_{i+1} \in R$",则将 w_{i-1}、w_i 以及 w_{i+1} 进行融合并赋予名词单元标签 NIU.名词单元融合示例如表3中示例7至10所示.

4) 状语单元融合规则

设标签集合 $P=\{$P, VV$\}$,标签集合 $Q=\{$NN, NR, LC, NIU$\}$,标签集合 $R=\{$LC$\}$.当文本中相邻的两个元素 w_i/t_i 与 w_{i+1}/t_{i+1} 的标签之间满足"$t_i \in P$ 且 $t_{i+1} \in Q$",则将 w_i 与 w_{i+1} 进行融合并赋予状语

单元标签 SIU；当文本中相邻的三个元素 w_{i-1}/t_{i-1}、w_i/t_i 以及 w_{i+1}/t_{i+1} 的标签之间满足“$t_{i-1}\in P$ 且 $t_i\in Q$ 且 $t_{i+1}\in R$”，则将 w_{i-1}、w_i 以及 w_{i+1} 进行融合并赋予状语单元标签 SIU. 状语单元融合示例如表3中示例11至12所示，其中 MIU，PEIU，NIU 及 SIU 为本文自定义标签.

表3 信息单元融合示例表

示例	融合前	融合后
3	开心/VV 的/DEG	开心的/MIU
4	张三/PERSON 的/DEG	张三的/MIU
5	一个/NIU 红色的/JJ	一个红色的/MIU
6	王某/PERSON 和/CC 李某/PERSON	王某和李某/PEIU
7	一/CD 个/M	一个/NIU
8	张三的/MIU 课本/NN	张三的课本/NIU
9	筷子/NN 和/CC 碗/NN	筷子和碗/NIU
10	这/DT 个/M 房间/NN	这个房间/NIU
11	在/P 硬盘/NN 里/LC	在硬盘里/SIU
12	坐/VV 椅子/NN 上/LC	坐椅子上/SIU

2.3 核心度序列

每个文本都有自己的核心内容，这一点在新闻语料中体现的更为明显，每篇新闻都有特定的核心内容，突出核心内容的方式主要是强调与重复. 在新闻语料中，强调的方式主要是在新闻标题和新闻正文首尾阐述新闻的核心内容，重复的方式主要是在叙述内容时核心内容相关的人、物都会被重复提及. 通过对大量语料的分析，可以发现新闻的核心内容所包含的信息单元比其信息单元作为事件成分的概率要大. 结合这个语言知识，通过计算核心度序列的方法来突出核心内容所包含的信息单元的重要性，当待抽取的分句中含有多个可以作为事件成分的信息单元时，优先抽取核心度高的信息单元.

在统计核心度的过程中，以概率统计为基础，利用标题、正文首尾句、正文等三类信息的加权值作为计算核心度的直接依据. 以统计新闻标题中信息单元的核心度为例，首先将新闻标题转换成 n 维向量空间模型 VSM，即 $(w_1,w_2,w_3,\cdots,w_n)=(c_1,c_2,c_3,\cdots,c_n)$，其中 w_i 表示在标题中出现的各个不同的信息单元，对应的 c_i 表示信息单元 w_i 在标题中出现的次数，根据向量空间模型中的值计算每个信息单元的核心度，信息单元 w_i 的核心度 p_i 计算方法如公式(1)所示.

$$p_i=c_i/\sum_{j=1}^{n}c_j \tag{1}$$

得到标题信息单元核心度序列 $P(p_1,p_2,\cdots,p_n)$ 之后，用同样的方法分别计算正文首尾句、正文中所有信息单元的核心度序列 $Q(q_1,q_2,\cdots,q_m)$ 与 $R(r_1,r_2,\cdots,r_k)$，特别地，在统计正文信息单元核心度序列时，不再重复统计正文首尾句中的信息单元. 最后将三个核心度序列进行整合，统一到一个序列 $L(l_1,l_2,\cdots,l_j)$ 中，其中 j 的值取决于标题、正文首尾句、正文中共含有多少个不同的信息单元，当一个信息单元同时存在于标题、正文首尾句、正文中的两者或三者时，将其核心度进行累加. 最终，得到的核心度序列中包含文本中所有信息单元的核心度.

2.4 原子事件抽取算法

本方法的最后一个步骤是原子事件抽取，算法1描述了原子事件抽取的过程. 为了提高原子事件抽取算法的效率，本文将原子句的基本句子结构分为三类，认为文本中出现的原子句句子结构均可以由这三类基本句子结构扩展联合得出. 这三类基本句子结构即一般句式、“把”字句式和“被”字句式，其中一般句式的基本结构为“主体＋事件谓词＋客体”，“把”字句式的基本结构为“主体＋把＋客体＋事件谓词”，“被”字句式的基本结构为“客体＋被＋主体＋事件谓词”.

算法1 原子事件抽取算法
名称：原子事件抽取算法
输入：信息单元融合后的文本
输出：原子事件集合
步骤0：初始化原子事件集合；
步骤1：导入核心度序列；
步骤2：将输入的信息单元融合后的文本切分为句子；
步骤3：for 每一个句子 do；
步骤4：抽取句子中的时空信息；
步骤5：将句子根据逗号切分为原子句；
步骤6：for 每一个原子句 do；
步骤7：for 每一个信息单元 do；
步骤8：根据核心度序列得到当前信息单元核心度；
步骤9：endfor；
步骤10：判断原子句句子结构；
步骤11：根据句子结构，抽取核心度最高的谓词作为事件谓词；
步骤12：if 事件谓词存在 do；

步骤13:根据句子结构,抽取核心度最高的信息单元作为事件参与者;
步骤14:if 事件参与者存在 do;
步骤15:将事件谓词、事件参与者和时空信息构成的原子事件添加到原子事件集合;
步骤16:endif;
步骤17:endif;
步骤18:endfor;
步骤19:endfor;

3　实验结果分析

3.1　实验语料

本文进行实验的语料集中的400篇新闻语料均来自于互联网,共包含142 131字,每篇新闻语料长度为200字至500字,平均长度355字.利用本文的方法对400篇新闻语料进行信息单元融合,然后手工标注其中的原子事件、原子事件谓词和原子事件参与者.经统计,400篇新闻语料中共有原子事件15 301组、原子事件谓词15 301个和原子事件参与者19 732个,平均每篇新闻语料含有原子事件38.2组、原子事件谓词38.2个和原子事件参与者49.3个.

3.2　评测方法

关于具体的原子事件、原子事件谓词、原子事件参与者评测方法,本文通过准确率(P)、召回率(R)和$F1$来进行度量,其具体计算方法如公式(2)、(3)和(4)所示.

$$P=\sum_{i=1}^{n}a_i/\sum_{j=1}^{n}b_j \tag{2}$$

$$R=\sum_{i=1}^{n}a_i/\sum_{k=1}^{n}c_k \tag{3}$$

$$F1=\frac{2\times P\times R}{P+R} \tag{4}$$

在计算原子事件、原子事件谓词、原子事件参与者的P、R、$F1$时,上式中的a_i表示每篇语料中正确抽取的原子事件、原子事件谓词、原子事件参与者的个数,b_j表示每篇语料中抽取的原子事件、原子事件谓词、原子事件参与者的个数,c_k表示每篇新闻语料中人工标注的原子事件、原子事件谓词、原子事件参与者的个数.在评测抽取的原子事件是否正确时,如果原子事件的事件谓词或者事件参与者抽取错误,则该原子事件记为错误抽取.在评测抽取的原子事件谓词是否正确时,如果该谓词不是分句中核心事件的谓词,则该事件谓词记为错误抽取.在评测抽取的原子事件参与者是否正确时,如果该参与者不是属于事件谓词所支配的参与者,则该原子事件参与者记为错误抽取.

3.3　实验结果

通过采用信息单元融合的事件抽取方法和未采用信息单元融合的事件抽取方法分别对实验语料进行事件抽取,原子事件谓词、原子事件参与者与原子事件的抽取结果如表4所示.

表4　原子事件谓词、原子事件参与者和原子事件的抽取结果　　%

参数	原子事件谓词		原子事件参与者		原子事件	
	未采用	采用	未采用	采用	未采用	采用
准确率	82.73	82.86	82.87	88.22	65.21	70.00
召回率	51.98	52.25	59.45	63.13	38.74	44.14
$F1$	63.84	64.09	69.23	74.24	48.60	54.14

从表4中的结果可以看出,采用信息单元融合的事件抽取方法的抽取结果比未采用信息单元融合的事件抽取方法的抽取结果是有较大提升的,尤其是在原子事件参与者抽取和原子事件整体抽取方面,原子事件参与者抽取结果的$F1$提升了5.01%,原子事件抽取结果的$F1$提升了5.54%.此外,采用信息单元融合的事件抽取方法之后,原子事件谓词的提升效果不是十分显著,原因在于原子事件谓词一般就是单个动词组成的,不需要过多的融合,而信息单元融合中与动词相关的信息单元融合规则也比较少,因此采用信息单元融合的事件抽取方法中原子事件谓词的抽取结果与未采用信息单元融合的事件抽取方法中原子事件谓词的抽取结果相比并没有很大提升.

原子事件抽取方法是由事件谓词驱动的,在抽取过程中首先抽取的是原子事件谓词,因此原子事件谓词的抽取效果对原子事件抽取以及原子事件参与者抽取都有影响.从表4可以看出,原子事件谓词抽取的准确率高达82.86%,而召回率只有52.25%,这是由于动词过滤环节所使用的词表不够充足,过滤的词类不够完善,导致在分句中出现嵌套事件或并列事件时,无法正确抽取原子事件谓词.经统计,在400篇新闻语料中,共有128个分句中含有嵌套事件,共有72个分句中含有并列事件,共有12个分句中既有嵌套事件又有并列事件.此外,当分句长度过长时,句子成分会十分复杂,例如,“他是前苏联援

华抗日志愿军武汉大会战战地故迹重游团的成员"的核心成分是"他是成员",但是由于"成员"一词的修饰成分太长,导致分词、词性标注过程无法正确拆分解析,在动词过滤阶段也很难将其中作为修饰的动词过滤掉,进而影响了信息单元融合过程,最终导致无法正确抽取句中的原子事件谓词.

从表4中可以看出,原子事件抽取效果劣于原子事件谓词的抽取效果,而原子事件参与者的抽取效果优于原子事件谓词的抽取效果,出现这种情况的原因,一方面是因为原子事件的评估过程既受事件谓词抽取结果的影响又受事件参与者抽取结果的影响,而原子事件参与者的评估是独立的;另一方面是因为一个原子事件往往涉及一个或者两个参与者,当原子事件涉及两个参与者时,只有当两个参与者同时抽取正确,该原子事件才记为正确抽取,因此,当只正确抽取两个参与者之中的一个参与者时,该原子事件记为错误抽取而参与者增加一个正确抽取.

4 结 论

本文采用基于信息单元融合的方法在新闻语料中进行原子事件抽取,这种方法是一种基于规则的方法,不需要大规模标注语料进行训练,简单实用,并且能够对各个领域的新闻语料进行事件抽取,不局限于某一类事件或者某一领域的事件,具有普遍适用性.下一步工作可以在原子事件抽取的基础上,进行事件之间关系的抽取,进而方便计算机理解整篇新闻的语义,而不仅仅局限于句义,还可以引入统计模型,利用统计模型与信息单元融合方法相结合的方法扩展规则库,提高原子事件抽取的效率.

参考文献:

[1] 赵妍妍,秦兵,车万翔,等.中文事件抽取技术研究[J].中文信息学报,2008,**22**(1):3-8.
Zhao Y Y, Qin B, Che W X, *et al*. Research on Chinese event extraction[J]. *Journal of Chinese Information Processing*, 2008, **22**(1):3-8(Ch).

[2] 王萌,李春贵,徐超,等.主题与子事件发现的多文档自动文摘[J].计算机工程与应用,2011,**47**(18):130-134.
Wang M, Li C G, Xu C, *et al*. Using topic and subevent discover to extract multi-document summarization[J]. *Computer Engineering and Applications*, 2011, **47**(18):130-134(Ch).

[3] Hakkani-Tur D, Ji H, Grishman R. Using Information Extraction to Improve Cross-lingual Document Retrieval[DB/OL]. [2014-03-05]. *http://www.lattice.cnrs.fr/poibeau/mmies/Proceedings_MMIES2007.pdf#page=24*.

[4] 肖升,何炎祥.基于动词论元结构的中文事件抽取方法[J].计算机科学,2012,**39**(5):161-164,176.
Xiao S, He Y X. Approach of Chinese event IE based on verb argument structure[J]. *Computer Science*, 2012, **39**(5):161-164,176(Ch).

[5] 侯立斌,李培峰,朱巧明,等.基于跨事件理论的缺失事件角色填充研究[J].计算机科学,2012,**39**(7):200-204.
Hou L B, Li P F, Zhu Q M, *et al*. Using cross-event inference to fill missing event argument[J]. *Computer Science*, 2012, **39**(7):200-204(Ch).

[6] Gupta P, Ji H. Predicting Unknown Time Arguments based on Cross-Event Propagation[DB/OL]. [2014-05-06]. *http://cs.nyu.edu/%7Ehengji/time.pdf*.

[7] Llorens H, Saquete E, Navarro-Colorado B. TimeML events recognition and classification learning CRF models with semantic roles[DB/OL]. [2014-03-06]. *http://aclweb.org/anthology-new/C/C10/C10-1082.pdf*.

[8] Ahn D. The Stages of Event Extraction[DB/OL]. [2014-04-02]. *http://dare.uva.nl/record/221799*.

[9] 徐红磊,陈锦秀,周昌乐,等.自动识别事件类别的中文事件抽取技术研究[J].心智与计算,2010,**4**(1):34-44.
Xu H L, Chen J X, Zhou C Y, *et al*. Research on event type identification for Chinese event extraction[J]. *Mind and Computation*, 2010, **4**(1): 34-44(Ch).

[10] 许旭阳,李弼程,张先飞,等.基于事件实例驱动的新闻文本事件抽取[J].计算机科学,2011,**38**(8):232-235.
Xu X Y, Li B C, Zhang X F, *et al*. News text event extraction driven by event sample[J]. *Computer Science*, 2011, **38**(8):232-235(Ch).

[11] 刘茂福,胡慧君.基于认知与计算的事件语义学研究[M].北京:科学出版社,2013:45-67.
Liu M F, Hu H J. *Event Semantics Based on Cognition and Computation*[M]. Beijing: Science Press, 2013: 45-67(Ch).

[12] Xia F. The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0)[EB/OL]. [2014-04-09]. *http://repository.upenn.edu/ircs_reports/38/*. 2000.

□

WUHAN DAXUE XUEBAO LIXUE BAN

中国科技核心期刊
北大中文核心期刊
美国化学文摘（CA）收录期刊
美国《剑桥科学文摘》（CSA）收录期刊
美国《数学评论》（MR）收录期刊
俄罗斯《文摘杂志》(P*）收录期刊
德国《数学文摘》（ZBI）收录期刊
中国期刊全文数据库收录期刊
中国科学论文引文数据库收录期刊

武汉大学学报（理学版）
双 月 刊　公开发行
1930年 创刊
第 61 卷　第 2 期（总第 270 期）
2015 年 4 月 24 日出版

Journal of Wuhan University
(Natural Science Edition)
Bimonthly,Started in 1930
Vol.61,No.2(Serial No.270)
Publishing on Apr. 24, 2015

主　　管　中华人民共和国教育部
主　　办　武 汉 大 学
主　　编　邓 子 新
执行副主编　谭　　辉
编　　辑　《武汉大学学报（理学版）》编辑部
（E-mail:whdz@whu.edu.cn 电话：027-68756952）
出　　版　武汉大学科学技术发展研究院
（湖北 武汉 珞珈山，邮政编码：430072）
http://whdy.cbpt.cnki.net
印 刷 装 订　武汉科源印刷设计有限公司
国 内 订 购　全国各地邮政局（所）
国 内 发 行　武汉市报刊发行局
国 外 发 行　中国国际图书贸易总公司
（北京399信箱，邮政编码：100044）

Supervisor : Ministry of Education,P. R. China
Sponsor : Wuhan University
Chief Editor : DENG Zixin
Deputy Chief Editor : TAN Hui
Editor : Editorial Department of Journal of Wuhan University (Natural Science Edition)
(E-mail:whdz@whu.edu.cn Tel:86-27-68756952)
Publisher : Research and Development Office of Wuhan University (Wuhan 430072,Hubei,China)
http://whdy.cbpt.cnki.net
Overseas Distributor : China International Book Trading Corporation
(P. O. Box 399,Beijing 100044,China)
Code No.BM312

刊号：ISSN 1671-8836 / CN 42-1674/N
代号：国内 38-8 / 国外 BM312
定价：国内 20.00元 / 全年 120.00元