

文章编号: 1003-0077(2014)03-0036-12

## 基于条件随机场的汉语框架语义角色自动标注

宋毅君<sup>1</sup>, 王瑞波<sup>1</sup>, 李济洪<sup>1</sup>, 李国臣<sup>2</sup>

(1. 山西大学 计算中心, 山西 太原 030006;  
2. 太原工业学院, 山西 太原 030008)

**摘要:** 在给定目标词及其所属框架的条件下, 汉语框架语义角色标注可以分为语义角色识别和角色分类两个步骤。该文将此任务通过 IOB2 标记策略形式化为词序列标注问题, 以词为基本标注单元, 采用条件随机场模型进行自动标注实验。先对语料使用清华大学的基本块自动分析器进行分析, 提取出 15 个块层面的新特征, 并将这些特征标记形式化到词序列上。以文献[20]已有的 12 个词层面特征以及 15 个块层面特征共同构成候选特征集, 采用正交表方法来选择模型的最优特征模板。在与文献[20]相同的语料上, 相同的 3 组 2 折交叉验证实验下, 语义角色标注的总性能的 F1-值比文献[20]的 F1-值提高了近 1%, 且在显著水平 0.05 的 t-检验下显著。实验结果表明: (1) 基于词序列模型, 新加入的 15 个块层面特征可以显著提高标注模型的性能, 但这类特征主要对角色分类有显著作用, 对角色识别作用不显著; (2) 基于词序列的标注模型显著好于以基本块为标注单元以及以句法成分为标注单元的标注模型。

**关键词:** 汉语框架语义知识库; 语义角色标注; 条件随机场模型; 基本块

**中图分类号:** TP391 **文献标识码:** A

### Semantic Role Labeling of Chinese FrameNet Based on Conditional Random Fields

SONG Yijun<sup>1</sup>, WANG Ruibo<sup>1</sup>, LI Jihong<sup>1</sup>, LI Guochen<sup>2</sup>

(1. Computing Center of Shanxi University, Taiyuan, Shanxi 030006, China;  
2. Taiyuan Institute of Technology, Taiyuan, Shanxi 030008, China)

**Abstract:** Given a predicate word and its frame, semantic role labeling of Chinese FrameNet can be divided into two steps: the boundary identification of semantic roles and the classification of semantic roles. In this paper, these tasks are formalized onto the word sequential labeling problem through IOB2 strategy. We apply conditional random field model to automatic labeling experiment with word as the basic tagging unit. We extract 15 new base-chunk features by applying the base chunk parser of Tsinghua University to automatic parsing on sentences, and the features are formalized onto the word sequence. Experiments show that the F1-value of the total performance of semantic roles labeling increases by nearly 1% in comparison with the baseline, which is significant under 0.05 significance level of the t-test.

**Key words:** Chinese FrameNet; semantic role labeling; conditional random fields; base chunk

## 1 引言

汉语框架语义知识库 CFN(Chinese FrameNet)<sup>[1]</sup>工程是以 C. J. Fillmore 的框架语义学<sup>[2]</sup>为理论基础, 以加州大学伯克利分校的英语 FrameNet 为参照, 以汉语真实语料为依据, 提供计算机使用的汉语

词汇语义知识库。目前汉语框架语义知识库中已手工构建框架 300 个左右, 20 000 条左右例句(每个例句已经分好词, 有词性标注、命名实体标注、目标词标注、语义角色标注、角色的短语类型和句法功能)。

框架语义学认为框架是跟一些激活性语境(motivating context)相一致的一个结构化的范畴系统, 是储存在人类经验中的图式化情境, 这种范畴

系统所描述的既可能是一个实体,也可能是一种行为实践模式,甚至是一些社会制度、习俗等。框架中的各种参与者称为框架元素(frame elements)。例如“波动、增加、提高、减少、降低”等汉语词语有共同的意义基础——都表示数量变化,归为一个框架(框架名“量变”)进行描述,以实体(Ent)、属性(Att)、初值(Val1)、终值(Val2)、初状态(Inis)、终状态(Finis)、变幅(Diff)、值区间(Val\_ran)等为框架元素进行刻画。框架元素在语义关系中的作用与语义角色或格角色相当,但传统的语义角色或格角色是相对于普遍的词汇而言的(像 PropBank 语料中所有定义的角色只有 50 多个),而框架元素仅适用于具有相同框架背景的一小组词语,其类型大大细化。框架元素表达的语义内容更加丰富、更加深入,用其来描述自然语言的语义更为适当,实用性强。在 CFN 工程中,句子标注是以框架库为基础,针对一个句子,给定一个词元和该词元所属框架,给框架元素所在的成分标记框架元素、短语类型和句法功能三种信息。例如,对“量变”框架的例句“大型弦乐队的人数增加一倍或一倍以上。”的标注结果如下:

<att-np-subj 大型弦乐队的人数> <tgt=[量变] 增加> <diff-np-obj 一倍或一倍 以上>。

其中,tgt 表示所标注的目标词“增加”,该词语属于[量变]框架;标注信息有三层,例如 att-np-subj 中,att 表示框架元素“属性”,np 表示短语类型是名词性短语,subj 表示句法功能是主语,其他标记依此类推。

根据框架语义学的理论,词语的语义理解是语句、篇章理解的关键。而词语的语义是由该词语所激起的框架所描述的场景来表示的,场景中的参与者就是该词语的语义角色。框架是人类认知中逐渐形成且固定下来的概念结构,这些概念结构相互联系形成网络。一个概念(框架)的理解不只是孤立考察这个概念本身,而是要将其放在整个框架网络中才能理解。一个句子乃至篇章的语义是可以由其中的词语激起的框架、这些框架之间的关系以及框架与句子中的成分(语义角色)之间的关系,来形式化表达的。这种表达是从更高的概念层次来表达句义,将为文本的理解提供非常重要的形式化语义信息。因此,识别词语所激起的框架及其语义角色标注是自动语义分析的首要环节。为此,近年来我们开始着手研究汉语框架语义角色的自动标注,并得到国家自然科学基金、国家 863 项目的支持。

在国际上,语义角色标注(Semantic Role Labeling, 简记为 SRL)任务的评测已举行过多次,我国

参评者在这些评测中取得很好的评测名次,如 CoNLL-2009 Shared Task 上车万翔等<sup>[3]</sup>、赵海等<sup>[4]</sup>。另外,基于评测使用的语料(英文 PropBank),刘挺<sup>[5]</sup>以句法成分为基本标注单元,使用最大熵模型研究了语义角色的自动标注问题,在开发集和测试集上分别获得了 75.49% 和 75.60% 的 F1-值。董静等<sup>[6]</sup>将句法分析树“压平”,并考虑句法树中水平层次上的角色标签之间的马尔科夫依赖关系,以线性链条件随机场为标注模型进行了语义角色标注实验,提高了模型的精度。于江德等<sup>[7]</sup>使用条件随机场模型研究了英文 PropBank 的语义角色标注问题,他将完全句法分析树转换成浅层短语块序列,并使用浅层短语块和命名实体块作为标注单位,也取得不错的结果。

汉语的语义角色自动标注的研究大部分是基于中文 PropBank 上的,主要工作有 Honglin Sun<sup>[8]</sup>、Yu<sup>[7]</sup>、Xue<sup>[9]</sup>、丁伟伟和常宝宝<sup>[10-12]</sup>、Weiwei Sun<sup>[13-14]</sup>、李军辉等<sup>[15]</sup>、李世奇等<sup>[16]</sup>。总的来说,绝大部分研究者的工作是通过构建新的特征来提高标注模型的性能。一般说来,语义角色标注任务主要分为角色识别和角色分类两个子任务。Xue<sup>[9]</sup>基于完全句法分析树,采用最大熵模型,在角色分类子任务(使用手工标记好的句法分析)得到了 94.1% 的 F1-值,但若采用自动的句法分析,却只有 71.9% 的 F1-值,结论是提高汉语的句法分析器是关键。丁伟伟和常宝宝<sup>[11-12]</sup>将语义角色标注任务建立在词的“semantic chunking”技术之上,不使用句法分析,采用线型链条件随机场模型,却得到了 72.64% 的 F1-值(好于 Xue<sup>[9]</sup>的结果 71.9%)。即使在角色分类子任务中,仅使用词层面特征也可以得到较高的标注性能,说明对汉语语义角色标注句法结构信息不是必不可少的。Weiwei Sun<sup>[13]</sup>基于一种新的组块分析方法,结合引进新的特征,取得了 76.46% 的 F1-值,而仅角色分类子任务也得到更好的结果<sup>[14]</sup>。

近年来,许多学者开展了基于英文 FrameNet 语料的语义角色标注的研究。Gildea 等人<sup>[17]</sup>使用概率插值建立模型,在整句的自动分析的完全句法分析树上提取特征,在 FrameNet 测试集上取得了 65% 的准确率和 61% 的召回率。在 Senseval-3 的 SRL Task 评测中,在给定目标词以及所属框架下,评测的最好结果达到 89.9% 的准确率和 77.2% 的召回率。SemEval 2007 Task-19 的任务除语义角色标注之外,增加了句子中框架的识别(包括目标词识

别、目标词所属框架名的排歧(类似 WSD))以及句子中所激活的框架的语义依赖关系(semantic dependency)的分析。只有 3 个队提交了评测结果,在 Task-19 给定的测试集上,只进行框架识别的 F1-值在 49%~75%之间;在自动识别框架的基础上,再进行语义角色标注的 F1-值在 36%~49%之间。

对汉语框架语义角色的自动标注,刘鸣洋等<sup>[18]</sup>基于规则的方法研究了“自主感知”和“非自主感知”两个框架的语义角色的标注,但每个框架构建规则工作量大,适应性差。刘开瑛等<sup>[19]</sup>采用统计学习模型对若干框架作了初步的自动标注研究。李济洪等<sup>[20]</sup>将角色标注任务看作词序列的 BIO 序列标注,以词为标注单元,提取了 12 个词层面若干特征(目标词、当前词、当前词词性、当前词的位置以及它们的组合特征和窗口特征),并给出了汉语框架语义角色标注的一种基于正交表的特征选择方法。对每个框架所给的语料,采用 3 组 2 折交叉验证实验,分别使用条件随机场模型建模。在由 25 个框架例句所构成的语料上,在给定的目标词及其所属框架下,最终,25 个框架语义角色标注总的 F1-值的宏平均达到 61.62%。这个实验结果主要是基于词序列而进行的标注,是以词为标注单元,不同于大多数文献中的以句法分析之后的句法成分为标注单元。

## 2 基于词序列上的汉语框架语义角色标注

从文献可以看出,大多数英语语料的语义角色标注都在句法分析之后,以句法成分为标注单元,中文是否也应该如此呢?

事实上,以 Stanford 大学的汉语句法分析器 v1.6,李济洪等<sup>[20]</sup>对实验数据集上所有例句进行了自动分析,并使用与 Xue<sup>[9]</sup>完全相同的特征在 25 个框架上进行同样的实验,即从自动完全句法树上提取 Xue<sup>[9]</sup>中完全相同特征,使用最大熵模型对句法分析树中的节点只进行角色识别,得到标注的结果只有 52.43%。这个结果明显低于文献<sup>[20]</sup>的结果。进一步分析发现,实验语料中,在自动句法分析后,角色的边界完全能够与句法分析树中的句法成分对齐的只有 76%<sup>[20]</sup>。汉语句法分析器的性能差是影响实验性能的主要原因。同样在 Xue<sup>[9]</sup>以中文 PropBank 为语料的语义角色标注研究中,也证实自动句法分析器的性能提高是关键。目前,汉语的句法分析器还很不成熟。

许多语言学家认为,汉语是一种意合语言,这种意合体现在字组成词,词组成短语,短语构成句子等,各层级句法的形式标记并不明显,因此,整个句子的完全句法分析需要分析句中所有成分之间的意义组合关系。而汉语的语义角色标注主要分析给定的目标词在句法语义上所支配的词语(角色)。因此,从自然语言处理角度讲,汉语的完整的句法分析似乎是比语义角色标注更难的任务。人们习惯地将语义角色标注放在句法分析之后,原因可能是因为英文或其他语言都这么做;其他自然语言使用的特征以及处理方法可以顺势借鉴过来。再者就是语言学上一般的研究路径也是从词法到句法,再到语义、语用。然而从这些年汉语的自然语言处理发展的成果来看,这样的思路似乎进展艰难。是否可以考虑从分词后直接到语义的处理?作者认为这是一条值得探索的途径。毋庸置疑,语义角色标注是语义处理的一个典型任务。

我们的思路是:汉语的分词技术相对比较成熟,在分词基础之上,将角色标注通过 IOB2 策略形式化到词序列上,并从自动句法分析器获取的各种句法标记信息中提取特征,将这些特征同样通过 IOB2 策略融入到词序列标注模型中。经过模型选择 and 数据分析,对角色标注有用的特征保留在模型中,没有用的则剔除。目前汉语的完全句法分析器性能差,离实用还有距离,但并不排除某些局部的信息对角色标注是有用的,比如,局部短语结构信息。关键是我们如何在语义角色标注模型中如何利用这些有用的信息。一般文献中是将角色标注全然建立在句法成分上,完全依赖句法分析的结果,导致标注不理想。而采用本文的思路,既有效利用了自动句法分析中部分有用的信息,充分发挥了统计模型的优点,又避免了在起初的标注环节就出现标注错误,即以句法成分为标注单元,自动分析的句法成分与语义角色的边界不能对齐的问题。

组块分析是一种浅层的句法分析。Surdeanu et al.<sup>[21]</sup>的英语的语义角色标注实验结果表明,完全句法分析和浅层句法分析的标注性能差异很小。Weiwei Sun<sup>[22]</sup>在中文 PropBank 上的实验也表明,基于浅层句法不比基于完全句法分析的结果差。重要的是组块分析可以将句子分割(若剔除功能词的话)在线性的块结构上,组块都位于同一个层次上,这使得我们容易将其标记信息形式化在词序列上,将角色标注转化为基于词序列结构的标注。

依据 Abney 的组块理论,结合汉语的特点,清

华大学周强<sup>[23]</sup>定义了汉语基本块的描述体系,并开发了一种自动组块分析器(下文称为基本块分析器)。所谓汉语的基本块主要描述句子中直接相邻的,以名词、动词、形容词等实词为中心聚合形成具有特定语义内容的词语序列,其中一般不包括各种功能词,如连词、叹词、语气词、助词、标点符号等。它们一般由 1~3 个词语组成,通过不同的外部句法表现和内部词汇关系形成各自特殊的概念内容描述体,成为汉语的字、词进入组块成句过程的基础和出发点<sup>[24]</sup>。基本块自动分析器是把一个输入的句子中的连接紧密的词组合生成一个个基本块,整个句子由基本块线性分割(若不考虑功能词的话)。在以词序列为结构的角色标注中,基本块的标记信息有可能对角色块内词的紧密联结提供有用信息。因此,在模型中充分利用自动标注的基本块信息,有可能提高角色标注的性能。

本文采用清华大学周强老师的基本块分析器<sup>[25]</sup>对选定语料做自动处理,提取了 15 种基于基本块的特征(下文中称为块层面特征)。在文献[20]中的 12 个词层面特征基础上,加入块层面特征,使用链条件随机场模型进行标注实验。实验结果表明块层面特征对角色自动标注有显著提高。

### 3 CFN 语义角色标注任务和词序列标注策略

一般来说,语义角色的标注任务为,给定一个句子以及句中目标词,自动标注该目标词所支配各个语义角色的边界(argument identification),自动标注角色的类别(argument classification)。

本文的 CFN 语义角色(框架元素)标注的任务定为:对给定的一个汉语句子,在给定目标词及其所属框架的前提下,自动识别语义角色的边界,自动标出该目标词所支配的语义角色。这个任务同 Senseval-3 的 SRL Task 相同。

在框架语义中,语义角色称为框架元素。框架元素包括核心框架元素、非核心框架元素。非核心框架元素中包括给定框架的非核心元素和通用非核心元素。通用非核心元素是所有框架都可共同拥有的非核心元素的集合,如时间(time)、空间(place)、工具(instru)等。为了与文献[20]的结果对比,本文同样暂不考虑通用非核心框架元素。

本文对所有例句使用 IOB2 策略,将 CFN 语义角色标注转化为词层面的线性序列标注问题。为此,将 CFN 语义角色自动标注步骤定为:1)角色

识别:自动识别出句子中语义角色块边界;2)角色分类:自动标注语义角色的类型;3)后处理:针对不合理标注串等明显标注错误进行后处理,输出合理的标注序列。根据 IOB2 策略中标记的不同,就有两种标注策略:第一种是将步骤 1)、2)同时做,即角色识别与角色分类一起进行,称为一步标注策略,其标注集合为{B-X, I-X, O},其中,X 为语义角色标记。第二种是每一步单独做,即先角色识别再角色分类,称为两步标注策略,这时角色识别的标记集合为{B, I, O},而角色分类的标记集合同一步标注策略。

例句:

<spkr 英方面> 作为报复措施也<tgt [陈述]宣布><msg 4 名俄罗斯大使馆的外交官为不受欢迎的人>。

使用 IOB2 标记将角色标注看作词序列的标注(一步标注策略),示例如下:

英|B-spkr 方面|I-spkr 作为|O 报复|O 措施|O 也|O 宣布|O 4|B-msg 名|I-msg 俄罗斯|I-msg 大使馆|I-msg 的|I-msg 外交官|I-msg 为|I-msg 不|I-msg 受|I-msg 欢迎|I-msg 的|I-msg 人|I-msg 。|O

其中,“宣布”是目标词,属于“陈述”框架。B-spkr 代表“陈述”框架中角色“说话者[spkr]”的开始,;I-spkr 表示“陈述”框架中“说话者[spkr]”角色的延续;O 表示块外词,即不是语义角色。

采用基本块的特征也是将相关特征信息表示到词层面,只不过这样需要在最后增加一步后处理,以便将自动标注出来的不合乎 IOB2 标记规则的序列处理成合法序列。本文的后处理算法与文献[20]中相同。

### 4 块层面特征的提取

本文所用的特征包括两部分,词层面特征和块层面特征。词层面特征是指,基于词、词性、词相对于目标词的位置,还有这些特征的组合而构成的特征。文献[20]表 1 共提取这样的特征 12 个。本文仍使用这些特征,这里所说的块层面特征,是指对语料中所有例句,使用自动基本块分析器分析之后,提取句法标记、结构标记信息而得到的特征。基本块分析器是把一个输入的句子中的连接紧密的词组合生成一个个基本块。整个句子由基本块线性分割(若不考虑功能词的话)。

本文使用了清华大学周强提供的基本块自动分析器。周强<sup>[24]</sup>定义的汉语基本块的描述体系中,每个基本块,使用句法标记和关系标记两种标记的组合,来描述它们的结构组合和语法关系。基本块标记集合,句法标记 8 种: np-名词块、mp-数量块、sp-空间块、tp-时间块、vp-动词块、ap-形容词块、dp-副词块、pp-介词块;关系标记 10 种: ZX-中心结构、LN-链式关联结构、LH-并列关系 CHC、PO-述宾关系 LCC、SB-述补关系 LCC、AD-附加关系 LCC、JB-介宾关系 LCC、SG-单词语块、RL-复杂关系、AM-歧义区间。

以本文第 3 节的例句为例,分析器输出为:

[np-AM 英方面] [vp-SG 作为] [vp-PO 报复措施] [dp-SG 也] [vp-SG 宣布] [mp-ZX 4 名] [np-ZX 俄罗斯大使馆] 的 [np-SG 外交官] [vp-AM 为不受欢迎] 的 [np-SG 人]。

由上面的例子可以看出,基本块分析器是将句子组合为比词大的基本块。所以,以词为基本标注单位需要将基本块信息通过 IOB2 策略表示到词上,再作为特征信息加入标注模型中。例如,基本块 [np-AM 英方面],通过 IOB2 策略,转化为句法标记特征: B-np 英 I-np 方面,关系标记特征: B-AM 英 I-AM 方面。

本文选用的 15 个特征详见表 1。为了区别于词层面的 12 个特征,特征编号从 13 到 27。下文中将这 15 个特征简称为块层面特征。

表 1 块层面特征列表及可选窗口大小

编号	特征	可选窗口大小		
13	句法标记	-	[0, 0]	[-1, 1] [-2, 2]
14	句法标记二元组	-	[-1, 1]	[-2, 2]
15	句法标记三元组	-	[-1, 1]	[-2, 2]
16	关系标记	-	[0, 0]	[-1, 1] [-2, 2]
17	关系标记二元组	-	[-1, 1]	[-2, 2]
18	中心词	-	[0, 0]	[-1, 1] [-2, 2]
19	中心词二元组	-	[-1, 1]	[-2, 2]
20	中心词性	-	[0, 0]	[-1, 1] [-2, 2]
21	中心词性二元组	-	[-1, 1]	[-2, 2]
22	句法与中心词搭配	-	[0, 0]	[-1, 1]
23	关系与中心词搭配	-	[0, 0]	[-1, 1]

续表

编号	特征	可选窗口大小		
24	浅层句法路径	-	[0, 0]	
25	命名实体	-	[0, 0]	
26	浅层子范畴	-	[0, 0]	
27	是否简单句	-	[0, 0]	

注:“-”表示不选取该特征,[-1,1]表示该特征开左右各为 1 大小的窗口,其他类推。可选窗口的大小基本上是根据实验者的经验判断。

下面以一个例句(见表 2)来对表 1 中特征作一些示范性说明:

(1) 命名实体的特征:当前标注单元是否是个命名实体。具体的命名实体标记需要通过词性的映射来决定。例如表 2 中第 6 列。

标注集合为:人名(PERSON),地名(LOCATION),时间名(TIME)和不是命名实体四类。其中,PERSON 包括的词性集合为{nh=人名、nhh=汉人名、nhf=汉姓氏、nhg=汉名字、nhy=音译名、nhr=日本人名、nhw=简称、笔名等},LOCATION 包括的词性集合为{ns=地名、nsh=中国地名、nssy=外国地名、ni=机构专名}。TIME 包括{nt=时间名词}。

(2) 句法标记特征:指当前标注的词所具有的句法标记,例如表 2 中第 2 列。

(3) 关系标记特征:指当前标注的词所具有的关系标记,例如表 2 中第 3 列。

(4) 块中心词特征:指当前标注的词所在基本块的中心词,例如表 2 中第 4 列。

(5) 块中心词的词性特征:指当前标注的词所在基本块的中心词的词性,例如表 2 中第 5 列。

(6) 浅层句法路径特征:指由当前所标注的词语和目标词之间的词性标记作为终结符,中间由基本块句法标记连接的链式序列,例如表 2 中最后一列。

(7) 浅层子范畴特征:由目标词的所在的基本块左边和右边基本块句法标记组成的序列,例如表 2 中倒数第 2 列。

(8) 是否为简单句:使用“是”和“否”来表示当前句子是否为简单句,例如表 2 中倒数第 3 列。

(9) 表中的组合特征与一般条件随机场模型中的组合特征一样。

表 2 块层面特征示例  
(表中第一行为表 2 对应的特征编号,第一列为词特征)

1	13	16	18	20	25	27	26	24
英	B-np	B-SG	英	jn	—	S	dp_vp_mp	(jn)np=>np=>vp=>vp=>dp=>vp(v)
方面	B-np	B-SG	方面	n	—	S	dp_vp_mp	(n)np=>vp=>vp=>dp=>vp(v)
作为	B-vp	B-SG	作为	v	—	S	dp_vp_mp	(v)vp=>vp=>dp=>vp(v)
报复	B-vp	B-PO	报复	v	—	S	dp_vp_mp	(v)vp=>dp=>vp(v)
措施	I-vp	I-PO	报复	v	—	S	dp_vp_mp	(n)vp=>dp=>vp(v)
也	B-dp	B-SG	也	d	—	S	dp_vp_mp	(d)dp=>vp(v)
宣布	B-vp	B-SG	宣布	v	—	S	dp_vp_mp	vp(v)
4	B-mp	B-ZX	名	q	—	S	dp_vp_mp	(m)mp<=vp(v)
名	I-mp	I-ZX	名	q	—	S	dp_vp_mp	(q)mp<=vp(v)
俄罗斯	B-np	B-ZX	大使馆	n	LOC	S	dp_vp_mp	(nsy)np<=mp<=vp(v)
大使馆	I-np	I-ZX	大使馆	n	—	S	dp_vp_mp	(n)np<=mp<=vp(v)
的	u	的	的	u	—	S	dp_vp_mp	u<=np<=mp<=vp(v)
外交官	B-np	B-SG	外交官	n	—	S	dp_vp_mp	(n)np<=u<=np<=mp<=vp(v)
为	B-vp	B-SG	为	v	—	S	dp_vp_mp	(v)vp<=np<=u<=np<=mp<=vp(v)
不	B-dp	B-SG	不	d	—	S	dp_vp_mp	(d)dp<=vp<=np<=u<=np<=mp<=vp(v)
受	B-vp	B-SG	受	v	—	S	dp_vp_mp	(v)vp<=dp<=vp<=np<=u <=np<=mp<=vp(v)
欢迎	B-vp	B-SG	欢迎	v	—	S	dp_vp_mp	(v)vp<=vp<=dp<=vp <=np<=u<=np<=mp<=vp(v)
的	u	的	的	u	—	S	dp_vp_mp	u<=vp<=vp<=dp<=vp<=np <=u<=np<=mp<=vp(v)
人	B-np	B-SG	人	n	—	S	dp_vp_mp	(n)np<=u<=vp<=vp<=dp<=vp<=np<=u <=np<=mp<=vp(v)
。	w	。	。	w	—	S	dp_vp_mp	w<=np<=u<=vp<=vp<=dp<=vp<=np<=u <=np<=mp<=vp(v)

注：表中块中心词特征(编号 18)，“报复措施”的中心词为“报复”，“4 名”的中心词为“名”，“俄罗斯大使馆”的中心词为“大使馆”。

## 5 基于正交表的特征模板选优方法

本文使用线型链条件随机场模型作为标注模型。在条件随机场模型中,特征选择实际上是特征模板的选择。模型的每个特征模板中不但包括特定的候选特征,还包括相应特征窗口的大小(由于每个候选特征均可以开适当大小的窗口)。显然,遍历所有可能的模板,再找出最优模板是 NP 难问题。对此,传统的做法是贪心算法,但贪心算法计算量较大,并且,由于每一个循环都依赖于上一步的结果,程序是顺序执行,因此,算法难以实现并行化。

文献[20]中,给了一种基于正交表实验的特征模板选优方法。文献[25]基于词层面的 12 个特征,将正交表方法与基于贪心算法的特征选择方法做了实验比较,贪心算法得到的 F1 值为 61.56%(标准差=0.00196),正交表得到的为 61.62%(标准差=0.0017),在显著水平 0.05 下,以异方差正态总体均值的 t-检验<sup>[26]</sup>,两者没有显著差异。但正交表方法实现的算法非常简单,且很容易实现并行处理,因此,特征模板的选择较容易。

将文献[20]中的 12 个词层面特征和本文新增的 15 个块层面特征,放到一起,共计 27 个特征,使用正交表  $L_{54}(2^1 \times 3^{25})$  来安排实验,这个正交表第 1

列为1个二水平列,其他25列均为三水平列,共有26个列。 $L_{54}(2^1 \times 3^{25})$ 安排特征如下:第12号特征“目标词”为必选特征,不用占用 $L_{54}(2^1 \times 3^{25})$ 中的列,第11号特征“词/目标词”只有两水平“[0,0]”,“-”,故应安排在表 $L_{54}(2^1 \times 3^{25})$ 中第一列,其他特征依次顺序安排在表中相应列。编号为24-27的特征,可选窗口大小只有2种,因此,在安排到 $L_{54}(2^1 \times 3^{25})$ 表的第23-26列时,将第三水平与第二水平设为相同,扩展为三水平。具体方案如表3所示。

以表3设置的水平给出的 $L_{54}(2^1 \times 3^{25})$ 正交实验方案可得到54个候选模板( $L_{54}(2^1 \times 3^{25})$ 的54个行),每个模板以3组2折交叉验证来实验,选择3x2交叉验证的平均F1-值最大者对应的模板为最优模板。

表3 全部特征列表及其可选窗口大小

编号	特征	可选窗口大小			$L_{54}(2^1 \times 3^{25})$ 中对应的列
		[0, 0]	[-1, 1]	[-2, 2]	
1	词	[0, 0]	[-1, 1]	[-2, 2]	2
2	词二元组	-	[-1, 1]	[-2, 2]	3
3	词性	[0, 0]	[-1, 1]	[-2, 2]	4
4	词性二元组	-	[-1, 1]	[-2, 2]	5
5	位置	[0, 0]	[-1, 1]	[-2, 2]	6
6	位置二元组	-	[-1, 1]	[-2, 2]	7
7	位置三元组	-	[-1, 1]	[-2, 2]	8
8	词与词性的搭配	-	[0, 0]	[-1, 1]	9
9	词与位置的搭配	-	[0, 0]	[-1, 1]	10
10	词性与位置的搭配	-	[0, 0]	[-1, 1]	11
11	词与目标词的搭配	-	[0, 0]		1
12	目标词(必选特征)	[0, 0]			
13	句法标记	[0, 0]	[-1, 1]	[-2, 2]	12
14	句法标记二元组	-	[-1, 1]	[-2, 2]	13
15	句法标记三元组	-	[-1, 1]	[-2, 2]	14
16	关系标记	[0, 0]	[-1, 1]	[-2, 2]	15
17	关系标记二元组	-	[-1, 1]	[-2, 2]	16
18	中心词	[0, 0]	[-1, 1]	[-2, 2]	17
19	中心词二元组	-	[-1, 1]	[-2, 2]	18
20	中心词性	[0, 0]	[-1, 1]	[-2, 2]	19
21	中心词性二元组	-	[-1, 1]	[-2, 2]	20
22	句法与中心词搭配	-	[0, 0]	[-1, 1]	21
23	关系与中心词搭配	-	[0, 0]	[-1, 1]	22

续表

编号	特征	可选窗口大小			$L_{54}(2^1 \times 3^{25})$ 中对应的列
		-	[0, 0]	[0, 0]	
24	浅层句法路径	-	[0, 0]	[0, 0]	23
25	命名实体	-	[0, 0]	[0, 0]	24
26	浅层子范畴	-	[0, 0]	[0, 0]	25
27	是否简单句	-	[0, 0]	[0, 0]	26

注:(1)候选特征的可选窗口大小最大选为三水平,这主要是为了适用于所选正交表 $L_{54}(2^1 \times 3^{25})$ 。具体调整的办法是,1到12号特征是文献[20]中表1去掉可选窗口的最后一列,并且对本表的7号特征(在文献[20]表1中对应为10号特征)的窗口大小做了微调;(2)表中每个特征的可选窗口大小是从表1中选来的,13、16、18、20号特征的可选窗口没有选“-”是作者认为这四个特征较为重要,必须在模型中出现。就像文献[20]表1中,词、词性、位置必须在模型中,而没有选“-”水平一样。

## 6 评价指标及显著性检验

### 6.1 评价指标

框架语义角色自动标注正确指的是语义角色的边界和类型均完全识别正确。在测试集上对每个框架的所有语义角色,采用信息检索中的准确率(precision)、召回率(recall)和F1-值来评价标注模型的性能。若使用k-fold交叉验证,以交叉验证实验的k个F1-值的平均值作为评价指标。若进行多次重复交叉验证实验,则再取平均的F1-值来最终评价标注模型的性能。

记:mP——表示多组交叉验证的平均准确率;  
mR——表示多组交叉验证的平均召回率;  
mF——表示多组交叉验证的平均F1-值;  
mA——表示多组交叉验证的平均精确率;  
Var(mF)——为mF的方差, std(mF)为mF的标准差。

### 6.2 语料的切分与3x2交叉验证

本文所用实验语料是从CFN框架库中选择例句个数较多的25个框架的例句库6692句,与文献[20]中所用例句以及语料的切分方法完全相同,详见文献[20]中表2。即将所选语料例句拆分为4份,其中每个框架的例句库按照词元均匀拆分到每一份中。本文采用2折交叉验证进行实验,即任取2份作为训练集,其他2份作为测试集,这样共可以做3组2折交叉验证(下文称为组块3x2交叉验证)。最终的评价指标以组块交叉验证实验的平均

F1-值来评价标注模型的性能。这样做的目的主要是要给出评价指标的较好的方差估计,以便从统计意义上判别不同结果的差别的显著性。

事实上,在评测中直接对各系统 F1-值大小进行排队、对比,而没有评价指标方差的估计,或者说没有统计意义上显著性分析,评测的排序结果并不具有严格的统计意义上的科学性,特别是在系统都采用统计机器学习建模的情况下。性能指标之间微弱的差异不能判定一个系统好于另一个系统,不能说明某种模型,或某种特征,或某种处理策略的有效性。即使在性能指标值之差看上去有较大的差别时,没有指标值方差的估计,也不能正确推断差异是否真实存在。因此,给出评价指标方差的估计以及相应的显著性检验方法至关重要。然而,有关这方面的研究文献中较少。

在非评测中,研究者一般是根据采用的语料自己划分训练、测试集,因此,不同论文中得到的性能指标值基本没有可比性,各模型的优劣无法正确判定。虽然,研究者在论文中在同样的语料上也对各种模型或各种处理策略进行对比,但语料划分不同,指标值的方差估计会有较大不同,因此,各模型的对比在没有指标值方差的正确估计下,同样难以得出可靠的结论。

一般来说,给定一个数据集(语料),使用机器学习方法来训练、测试主要有三个目的:1)分析某种特征是否能显著提高系统性能,或某种处理是否对性能指标的提高有用,这称为对性能指标有显著性影响的因素分析;2)选择一个泛化能力较好的模型,并同时给出模型的泛化能力的估计,称为模型选择以及泛化能力的估计;3)得到一个在开发软件系统中使用的模型的参数估计(实用模型)。

交叉验证是一种有效的模型选择以及模型泛化能力估计方法。在自然语言处理领域中,广泛使用交叉验证来做模型选择。那么如何在用交叉验证选模型时也同时给出性能指标的方差的估计?交叉验证在用作模型选择时其折数如何确定?

考虑到自然语言处理中是高维数据,模型训练、测试的计算量都比较大,因此,在使用交叉验证时,折数一般不能选得太大。而当语料规模较小时,还要考虑到训练、测试集的划分,尽量使得训练集与测试集的差异较小,因为这样才能基本满足统计机器学习中的同分布<sup>[26-28]</sup>的基本假设。就交叉验证方法的特性而言,相对较大的训练集,较小的测试集容易选择含特征较多的模型,导致模型过拟合,且选择

的一致性(consistent)<sup>[29]</sup>差;而相对较小的训练集,较大的测试集容易导致模型的泛化误差估计偏差大。因此,平衡训练集和测试集,使其大小基本相同是交叉验证模型选择<sup>[29]</sup>中重要的设置(尤其是在自然语言处理中)。同时,为了得到评价指标方差的较好的估计,可以考虑采用有一定的次数的重复训练、测试。

基于上述的思想,李济洪等<sup>[20]</sup>提出了组块 3×2 交叉验证方法,下文基于此给出了平均 F1-值的方差估计以及两个性能指标差异的显著性检验方法,用于性能指标的影响因素分析。

### 6.3 评价指标的方差的估计

根据上述语料的数据集切分方法,分成了 3 组。针对其中的每一组,使用了 2 折交叉验证,共有 6 个实验结果,表示成下表:

表 4 组块 3x2 交叉验证的相关指标

组别	训练集 ← …… → 测试集		评价指标		评价指标的平均值
			CV <sub>1</sub>	CV <sub>2</sub>	
1	D <sub>1</sub> <sup>(1)</sup>	D <sub>2</sub> <sup>(1)</sup>	F <sub>1</sub> <sup>(1)</sup>	F <sub>2</sub> <sup>(1)</sup>	F <sup>(1)</sup>
2	D <sub>1</sub> <sup>(2)</sup>	D <sub>2</sub> <sup>(2)</sup>	F <sub>1</sub> <sup>(2)</sup>	F <sub>2</sub> <sup>(2)</sup>	F <sup>(2)</sup>
3	D <sub>1</sub> <sup>(3)</sup>	D <sub>2</sub> <sup>(3)</sup>	F <sub>1</sub> <sup>(3)</sup>	F <sub>2</sub> <sup>(3)</sup>	F <sup>(3)</sup>

其中,D 代表切分后的数据集,上标为组别,下标为交叉验证索引号。

记  $F_i^{(k)}$  为每第  $k$  组中第  $i$  份交叉验证实验结果的 F1-值。记  $F^{(k)}$  为第  $k$  组中两份交叉验证实验结果 F1-值的均值。即:

$$F^{(k)} = \frac{1}{2} \sum_{j=1}^2 F_j^{(k)}$$

记 3 组交叉验证 F1-值的总平均为:

$$mF = \frac{1}{3} \sum_{i=1}^3 F^{(i)} = \frac{1}{6} \sum_{i=1}^3 \sum_{j=1}^2 F_j^{(i)}$$

文献[25]给出了  $\text{Var}(mF)$  的估计为:

$$\hat{\text{Var}}(mF) = \frac{1}{6} \sum_{i=1}^3 \sum_{j=1}^2 (F_j^{(i)} - F^{(i)})^2$$

注意,这里虽然表达式前面系数为 1/6,但不同于用 6 个 F1-值直接得到的样本方差,这里平方里面用的是组内差。

本文在以下部分为了书写方便,直接用  $\text{Var}(mF)$  来代表  $\text{Var}(mF)$  的估计。

本文采用统计学中两个方差未知且不相等的正态总体均值标准检验方法<sup>[30]</sup>来作为两评价指标差异的显著性检验。为了读者阅读方便,使用容易懂



的专门记号  $p\_value(mF_A, mF_B)$  来代表系统 A, B 的性能指标  $mF_A, mF_B$  的显著的显著性检验的  $p$  值。即

$$p\_value(mF_A, mF_B) = P\left\{\frac{|mF_A - mF_B|}{\sqrt{\text{Var}(mF_A) + \text{Var}(mF_B)}} \geq t_{1-\alpha/2}(n)\right\}$$

本文显著水平  $\alpha$  取值 0.05, 当  $p\_value(mF_A, mF_B) \leq 0.05$  时, 称两系统的性能有显著差异。

## 7 实验结果及分析

以下实验中所用的基本块分析器是由清华大学周强老师提供的, 在开放测试语料下, 其性能指标 F1-值(边界识别+成分标注)达到 89.47%<sup>[25]</sup>。这里的成分标注指的是句法标注+关系标注。本文所有实验所用 CRF++ 工具包来自于 Sourceforge<sup>[31]</sup>, 采用了其中的 CRF L2 算法, 使用了输出序列的一阶转移特征, 并选取调节参数  $C=1$  进行参数平滑。

实验一: 探索加入 15 个块层面特征是否能显著提高标注性能

特征选用: 将文献[20]中的 12 个词层面特征和本文新增的 15 个块层面特征, 放到一起, 共计 27 个特征, 使用正交表  $L_{54}(2^1 \times 3^{25})$  表 3 来安排实验。

标注步骤: 角色识别与角色分类同时标注(一步

标注策略)

标注模型: 25 个框架, 每个框架构建一个条件随机场模型

表 5 两种正交表方案的 25 个框架总实验结果对比

选用的特征	mP/%	mR/%	mF/%	std/mF
12 个词层面特征(文献[8]的结果)	74.16	52.70	61.62	0.0017
12 个词层面+15 个块层面特征	75.16	53.63	62.60	0.0006

实验结果分析:

由于  $p\_value(61.62\%, 62.60\%) < 0.05$ , 故两者有显著差异。说明加入基本块特征能显著提高性能。

实验二: 进一步深入分析块层面特征分别对角色识别和角色分类的作用

特征选用: 12 个词层面特征+15 个块层面特征

标注步骤: 先角色识别, 再角色分类(两步标注策略)

标注模型: 对角色识别, 所有框架一个模型, 而对角色分类, 每个框架构建不同的模型

正交表选用: 选择正交表  $L_{54}(2^1 \times 3^{25})$  来进行实验, 实验安排同表 3。

表 6 加入块层面特征的 CRF 模型的实验结果

特征选用	标注策略	mP/%	mR/%	mF/%	std/mF	mA/%	std/mF
12 个词层面特征	只角色识别	73.21	64.38	68.51	0.0057		
	只角色分类					83.05	0.0031
	先角色识别, 再分类	63.04	60.20	61.58	0.0049		
12 个词层面特征+15 个块层面特征	只角色识别	73.18	64.43	68.52	0.0035		
	只角色分类					84.08	0.0022
	先角色识别, 再分类	64.28	61.03	62.61	0.0044		

实验结果分析:

(1) 仅从角色识别来看, 基于词层面特征的结果为 68.51% ( $\sigma=0.0057$ ), 而加入块层面特征后为 68.52% ( $\sigma=0.0035$ ), 而  $p\_value(68.51\%, 68.52\%) > 0.8$ , 两者无显著差异。说明加入块层面特征对角色边界识别无显著作用。

(2) 在已知角色的边界情形下, 加入块层面特征后, 角色分类结果为 84.08% ( $\sigma=0.0022$ ), 基于词层面特征的结果为 83.05% ( $\sigma=0.0031$ ), 且  $p\_value(84.08\%, 83.05\%) < 0.05$ , 故有显著差异。

说明, 加入块层面特征, 主要是对角色分类性能提高有作用。这可能是因为基本块特征主要描述块内的句法结构和关系结构, 角色的类别与这些特征有关。

(3) 与角色识别和角色分类同时标注(一步标注策略)的实验一的结果比较  $p\_value(62.61\%, 62.60\%) > 0.8$ , 两者无差异。说明分开两步做没有提高性能。但可以明显看出, 两步分开做的评价指标的方差比两步一起做的方差大。这可能是由于

各框架的角色不同,在边界识别时,虽然 IOB2 标记相同,可以放在一起训练一个模型,但毕竟角色涵义差异较大,导致结果波动较大。

### 实验三：以基本块为标注单位的实验

在我们的实验中,在使用自动基本块分析器之后,角色边界与基本块边界不能对齐的只有大约 2%,因此,我们也进行了以基本块为标注单位的实验。这里的做法是,以基本块的中心词(可由基本块分析自动分析得到)代替基本块<sup>[32]</sup>,将原句缩减,将缩减处理后的语料作为训练、测试语料。形式上类似词序列标注。例如,本文第 3 节的例句,在本文第 4 节自动基本块标注后,缩减为:

英|B-spk 作为|O 报复|O 也|O 宣布|O 名|B-msg 大使馆|I-msg 的|I-msg 外交官|I-msg 为|I-msg 不|I-msg 受|I-msg 欢迎|I-msg 的|I-msg 人|I-msg 。|O

特征选用:因为以基本块为标注单元,这时词、词性特征就没有必要了,因此,以表 3 中去掉第 1 到 4 号、8 到 11 号特征之后所剩的 19 个特征为可选特征。

标注步骤:角色识别与角色分类同时标注(一步标注策略)

标注模型:25 个框架,每个框架构建一个条件随机场模型

正交表选用:仍然选择正交表  $L_{54}(2^1 \times 3^{25})$ ,正交表中列设置为:第 1 到 4 号、8 到 11 号特征空下,其他列同实验一。

表 7 以基本块为标注单元的 25 个框架总实验结果与其他实验的对比

选用的特征	mP/%	mR/%	mF/%	std/mF
12 个词层面 + 15 个块层面特征	75.16	53.63	62.60	0.000 6
用 Stanford 大学句法分析器,再用 ME 分类器,使用 Xue (2008)的特征	70.46	41.75	52.43	0.005 3
以基本块为标注单位,19 个特征	73.13	52.84	61.35	0.002 8

### 实验结果分析

(1)  $p\_value(62.60\%, 61.35\%) < 0.05$ ,  $p\_value(62.60\%, 52.43\%) < 0.05$ , 两者有显著差异。说明基于词序列的加入块层面特征的模型标注结果显著好于基于块为标注单元和基于句法成分为标注

单元的结果。

(2)  $p\_value(61.35\%, 52.43\%) < 0.05$ , 说明基于块为标注单元的结果显著好于基于句法成分为标注单元的结果。

## 8 结论

本文将汉语框架语义角色的自动标注问题形式化为词序列标注问题,并将自动分析的组块层面的标记信息形式化到词序列上来构建新特征,使用条件随机场模型研究了汉语框架语义角色的角色识别和角色分类任务。本文使用的组块分析器是清华大学周强的基本块分析器<sup>[25]</sup>。在自动基本块分析之后,新提取了 15 个块层面的特征,在文献[20]的 12 个词层面特征基础上,选用正交表  $L_{54}(2^1 \times 3^{25})$  为特征模板的选择方案。在与文献[20]相同的语料上,相同的组块  $3 \times 2$  交叉验证实验下,在给定句中的目标词及其所属框架下,25 个框架语义角色标注总的 F1 值的宏平均比文献[20]的 F1 值提高了近 1%。在显著水平 0.05 的异方差正态总体均值的 t-检验下,是显著的。实验结果表明:

(1) 加入 15 个块层面特征可以显著提高标注模型的性能,但这类特征主要对角色分类有显著作用,对角色的边界识别作用不显著。这可能是因为基本块特征主要描述块内的句法结构和关系结构,角色的类别与这些特征有关。

(2) 每个框架训练一个模型,角色识别与角色分类一起进行,与先角色识别,再角色分类两个步骤在标注性能上没有显著差别,但由前者得到的 25 个框架的整体的标注性能指标有较小的方差。

(3) 基于词序列的加入块层面特征的模型标注结果显著好于基于块为标注单元和基于句法成分为标注单元的结果。因此,基于词序列是一个可取的语义角色标注的出发点。

(4) 在全部 25 个框架的所有实验中,语义角色识别最好的结果(mF)为 68.52%;在给定语义角色边界下角色的最好结果(mA)为 84.08%;在给定句子中的目标词以及目标词所属的框架情况下,语义角色标注整个任务的最好结果(mF)达到 62.61%。就角色识别、角色分类两个子任务来说,角色识别是影响标注性能提高的主要步骤。

### 致谢

本文的实验中使用了清华大学周强老师的基本

块自动分析器,在此表示感谢。

## 参考文献

- [1] You L, Liu K. Building Chinese FrameNet Database [A]. Proceedings of IEEE NLP-KE'05[C]. Wuhan: IEEE, 2005: 301-306.
- [2] Fillmore, Charles J. Frame semantics and the nature of language[A]. In Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech[C]. 1976, 280: 20-32
- [3] Che WX, Li ZH, Li YQ, et al. Multilingual dependency-based syntactic and semantic parsing[A]. Proceedings of the CoNLL-2009 [C], Boulder: ACL Press, 2009: 49-54.
- [4] Zhao H, Chen WL, Kit C, Zhou GD. Multilingual dependency learning: A huge feature engineering method to semantic dependency parsing[A]. Proceedings of the CoNLL-2009[C]. Boulder: ACL Press, 2009: 55-60.
- [5] 刘挺,车万翔,李生. 基于最大熵分类器的语义角色标注[J]. 软件学报. 2007, 18(3):565-573.
- [6] 董静,孙乐,吕元华,冯元勇. 基于线性链条件随机场模型的语义角色标注[A]. 中国中文信息学会二十五周年学术会议[C]. 2006.
- [7] Yu JD, Fan X, Pang W, Yu Z. Semantic role labeling based on conditional random fields [A]. Journal of Southeast University(English Edition). 2007, 23(3): 361-364.
- [8] Sun HL, Jurafsky D. Shallow Semantic Parsing of Chinese[A]. Proceedings of NAACL-HLT 2004 [C]. 2004.
- [9] Xue NianWen. Labeling Chinese predicates with semantic roles[J]. Computational Linguistics, 2008, 34(2):225-255.
- [10] 丁伟伟,常宝宝. 基于最大熵原则的汉语语义角色分类[J]. 中文信息学报. 2008, 22(6):20-27.
- [11] Weiwei Ding, Baobao Chang. Fast Semantic Role Labeling for Chinese Based on Semantic Chunking[A]. Proceedings of 22nd International Conference on the Computer Processing of Oriental Languages (ICCPOL 2009)[C]. Hongkong, China. 2009.
- [12] Weiwei Ding, Baobao Chang. Word Based Chinese Semantic Role Labeling with Semantic Chunking [J]. International Journal of Computational Processing Oriental Language. 2009, 22(2-3): 133-154.
- [13] WeiWei Sun. Semantics-driven shallow parsing for Chinese semantic role labeling [A]. Proceedings of the ACL 2010[C]. 2010.
- [14] WeiWei Sun. Improving Chinese semantic role labeling with rich syntactic features[A]. Proceedings of the ACL 2010 Conference [C]. Uppsala, Sweden. 2010: 168-172.
- [15] 李军辉,周国栋,朱巧明,钱培德. 中文名词性谓词语义角色标注[J]. 软件学报, 2011, 22(8):1725-1737.
- [16] 李世奇,赵铁军,李晗静,刘鹏远,刘水. 基于特征组合的中文语义角色标注[J]. 软件学报, 2011, 22(2): 222-232.
- [17] Gildea D, Jurafsky D. Automatic labeling of semantic roles[J]. Computational Linguistics, 2002, 28(3): 245-288.
- [18] 刘鸣洋,由丽萍. 汉语感知词语的语义角色标注规则初探[A]. 内容计算的研究与应用前沿(CNCCL-2007)[C]. 北京:清华大学出版社. 2007,320-325,
- [19] 刘开瑛. 汉语框架语义网(CFN)构建现状[A]. 第四届全国学生计算语言学研讨会论文集[C]. 2008.
- [20] 李济洪,王瑞波,王蔚林,李国臣. 汉语框架语义角色自动标注[J]. 软件学报, 2010,30(4): 597-611.
- [21] Surdeanu M, Marquez L, Carreras X, Comas PR. Combination strategies for semantic role labeling[J]. Journal of Artificial Intelligence Research, 2007, 29: 105-151.
- [22] Weiwei Sun, Zhifang Sui, Meng Wang, and Xin Wang. Chinese semantic role labeling with shallow parsing[A]. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing [C]. Singapore: Association for Computational Linguistics. 2009,1475-1483.
- [23] 周强. 基于规则的汉语基本块自动分析器[A]. 第七届中文信息处理国际会议论文集[C]. 北京:电子工业出版社. 2007: 137-142.
- [24] 周强. 汉语基本块描述体系[J]. 中文信息学报, 2007, 21(3):21-27.
- [25] 李济洪. 汉语框架语义角色自动标注技术研究[D]. 山西大学 2010 届博士论文.
- [26] Ethem Alpaydin. Combined 5 x 2 cv F test for comparing supervised classification learning algorithms [J]. Neural Computation. 1999, 11(8):1885-1892.
- [27] Markatou M, Tian H, Biswas S, et al. Analysis of variance of cross-validation estimators of the generalization error. Journal of Machine learning Research, 2005, 6:1127-1168.
- [28] Sylvain Arlot. A survey of cross-validation procedures for model selection [J]. Statistics Surveys. 2010, 4:40-79. DOI: 10.1214/09-SS054.
- [29] YuHong Yang. Comparing Learning Methods for Classification[J]. Statistica Sinica. 2006, 16(2): 635-657.
- [30] 茆诗松. 统计手册[M]. 北京:科学出版社,2003.
- [31] Taku Kudo. CRF++ Tools Package: <http://crfpp.sourceforge.net/>, version:5.0, 2007.

[32] 王瑞波. 基于条件随机场的汉语框架语义角色自动

标注研究[D]. 山西大学 2009 届硕士论文.



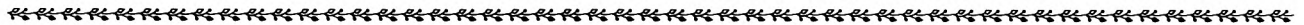
宋毅君(1977—), 硕士, 讲师, 主要研究领域为自然语言处理、机器学习。  
E-mail: songyj@sxu.edu.cn



王瑞波(1985—), 硕士, 助理实验师, 主要研究领域为统计自然语言处理。  
E-mail: wangruibo@sxu.edu.cn



李济洪(1964—), 博士, 教授, 主要研究领域为统计自然语言处理和机器学习。  
E-mail: lijh@sxu.edu.cn



## 第十一届全国自然语言处理青年学者研讨会在乐山师范学院召开

2014 年 5 月 15—16 日, 由中国中文信息学会主办, 乐山师范学院承办的“第十一届全国自然语言处理青年学者研讨会”(YSSNLP2014)在四川乐山举行。

开幕式上中国中文信息学会副理事长兼秘书长孙乐研究员, 中国中文信息学会副理事长、“973”首席科学家孙茂松教授分别代表学会致辞。乐山师范学院校长胡丹教授致辞, 并代表学校对各位专家、学者到校参会表示热烈欢迎。计算机科学学院院长金澎博士主持开幕式。

研讨会邀请到微软亚洲研究院主任研究员林钦佑博士、中科院计算所研究员兼爱尔兰都柏林城市大学教授刘群博士分别做大会特邀报告。会议安排了 14 个口头报告, 23 个技术海报, 同时还围绕“深度学习”和“众包”两个主题进行了两场主题讨论。国内自然语言处理领域的青年学者, 包括清华、北大等高校, 中科院计算所、自动化所等科研机构, 百度、微软等企业均派代表参加了此次研讨会。会议闭幕式由学会青年工作委员会主任、百度高级研究员赵世奇博士主持。在闭幕式上, 下届会议的承办单位代表、福州大学廖祥文博士对 YSSNLP2015 的承办工作做了简单介绍。经过激烈的角逐, 会议最后确定 YSSNLP2016 会议的承办单位为西北师范大学。

该会议自 2004 年创立以来已经逐渐成为国内自然语言处理领域产学研的最重要盛会之一, 旨在促进自然语言处理领域国内外学者间的学术互动, 加强研究人员与企业界和政府相关部门的交流对话, 共同促进整个自然语言处理产业的提升。前十届 YSSNLP 会议分别是在上海(复旦大学)、哈尔滨(哈尔滨工业大学)、厦门(厦门大学)、苏州(苏州大学)、北京(北京语言大学)、大连(大连理工大学)、沈阳(东北大学)、昆明(昆明理工大学)、深圳(哈工大深圳研究生院)和长沙(国防科学技术大学)召开。

# 基于条件随机场的汉语框架语义角色自动标注



作者: [宋毅君](#), [王瑞波](#), [李济洪](#), [李国臣](#), [SONG Yijun](#), [WANG Ruibo](#), [LI Jihong](#), [LI Guochen](#)  
作者单位: [宋毅君, 王瑞波, 李济洪, SONG Yijun, WANG Ruibo, LI Jihong\(山西大学计算中心, 山西太原, 030006\)](#), [李国臣, LI Guochen\(太原工业学院, 山西太原, 030008\)](#)  
刊名: [中文信息学报](#)   
英文刊名: [Journal of Chinese Information Processing](#)  
年, 卷(期): 2014, 28(3)

引用本文格式: [宋毅君. 王瑞波. 李济洪. 李国臣. SONG Yijun. WANG Ruibo. LI Jihong. LI Guochen 基于条件随机场的汉语框架语义角色自动标注\[期刊论文\]-中文信息学报 2014\(3\)](#)