

文章编号: 1003-0077(2008)06-0020-07

基于最大熵原则的汉语语义角色分类

丁伟伟, 常宝宝

(北京大学 计算语言学研究所, 北京 100871)

摘要: 语义角色标注是近些年来兴起的自然语言处理的一个新的研究领域。与英语方面的研究相比, 汉语方面的工作还不是很充分。该文在参考已有工作的基础上, 基于最大熵原则, 对汉语语义角色标注中的一个方面——语义角色分类进行了深入的研究。在提出了一些新的特征之后, 该文还充分利用了语义角色之间的相关性, 提取语义角色的上下文特征, 从而提高标记的准确率; 此外, 通过对不同特征的单独研究, 笔者发现了不同特征取得最优值时的窗口大小差别很大。发现这一现象后, 笔者设计了一种基于贪心策略的选择算法, 对不同的特征选择不同的窗口大小, 使得标记结果进一步提高。在综合采用了以上的策略之后, 笔者的汉语语义角色分类系统可以达到 95.00% 的准确率, 比前人有着较为显著的提升。从而证明了笔者的方法是有效的。

关键词: 计算机应用; 中文信息处理; 语义角色分类; 最大熵; 特征; 上下文; 窗口; 贪心策略

中图分类号: TP391

文献标识码: A

Maximum Entropy-Based Semantic Role Classification

DING Wei-wei, CHANG Bao-bao

(Institute of Computational Linguistics, Peking University, Beijing 100871, China)

Abstract The semantic role labeling (SRL) is a new research area of natural language processing in recent years. Compared to the study in English, Chinese SRL is still in its infancy stage. In this paper, we focus on the semantic role classification (SRC), one key step of SRL. Besides introducing some new features, we also explore the interdependence of the semantic roles. We employ the context features to improve the performance of the semantic role classification. And a greedy algorithm is designed to select the different windows of the context for different feature templates, since the highest performance can be achieved with different window sizes for different feature templates. In the experiments, the precision of our SRC system can achieve 95.00%, proving the validness of our approach.

Key words: computer application; Chinese information processing; semantic role classification; maximum entropy; features; context; window; greedy-strategy

1 引言

语义角色标注(Semantic Role Labeling), 又称浅层语义分析(Shallow Semantic Parsing), 指的是分析句子的论元结构, 即标记出句子中某个动词的所有论元, 属于语义分析任务。之所以称为浅层语义分析, 是因为该任务并不涉及深层次的语义分析和计算。

语义角色标注起始于 Dan Gildea 和 Dan Jurafsky^[1], 他们的实验所用语料是 Berkeley 大学开发的 FrameNet^[2]。在此之后, 语义角色标注这个任务逐渐得到了国际的关注, 这体现为两个方面: 一方面是语料的不断丰富。在 FrameNet 之后, 宾州大学在树库的基础上完成了英文 PropBank^[3], 并有与之相关的 VerbNet^[4] 等配套语义词典的构建; 另一方面是出现了一些相关的国际评测。CoNLL 2004^[5] 和 CoNLL 2005^[6] 都包含了语义角色标注的

收稿日期: 2008-06-05 定稿日期: 2008-09-02

基金项目: 国家自然科学基金资助项目(60303003); 国家社会科学基金资助项目(06BYY048)

作者简介: 丁伟伟(1985—), 男, 硕士生, 主要研究方向为自然语言处理; 常宝宝(1971—), 男, 副教授, 主要研究方向为自然语言处理。

任务。每次都有多家单位的参加,他们方法各不相同,对这个问题有着比较好的研究。语料的构建和评测的出现使语义角色标注的研究呈现出蓬勃的发展势头。依据研究方法的不同,大致可以分为两条道路,一个是以 Pradhan 等^[7]为代表的基于特征的方法,另一个是以 Moschiti 等^[8],Zhang 等^[9]为代表的基于核函数的方法。

相比较英语上的研究,中文语义角色标注的工作既开展比较晚,研究得也不是很充分。最早进行研究的是 Sun 等^[12],由于在当时还没有中文方面的专门语料,所以他们只是人工标记了包含某些动词的一些语料,并在这些语料上进行研究。虽不成系统,但是毕竟是一个有意义的开端。后来,伴随着中文 PropBank^[13]的构建, Xue Nianwen 开始了比较系统的中文语义角色标注的工作,并得出了一些很有意思的结论,比如:语义角色识别和语义角色分类所采用的特征是有区别的。这些工作不仅对中文的语义角色标注很有意义,也对英文的语义角色标注有所启发。

国内对汉语语义标注的研究最早起始于刘挺等^[10],于江德^[11]等,不过他们的研究重点仍然集中在英文的语义角色标注之上,实验的语料是 CoN LL-2005 的评测语料。刘怀军等^[14]针对汉语进行了语义角色标注的研究工作,目前也还主要局限在语义角色分类,没有一个公开的完整的语义角色标注系统。此外还有吕德新等^[15],他们的研究集中在特定句式,系统考察了疑问句句点论元的标注。不过总的来说,与英文上的工作相比,汉语语义角色标注方面的研究比较少,相关的文章也不是很多。

在目前的研究中,语义角色标注任务通常由两个阶段组成:前一个阶段是挑选出句法树上可能充当动词论元成分,这是语义角色识别。后一个阶段的任务是对识别出来的节点进行分类,具体判断出是指定动词的哪类论元,这是语义角色分类。这样的区分使得语义角色标注任务得到合理分解,处理的复杂性得到一定的简化。

本文的研究集中在汉语语义角色标注第二个阶段——汉语语义角色分类。工作建立在句法分析和语义角色识别结果正确的基础上,这样的做法与前人相同。本文组织方式如下:第二节是对中文 Proposition Bank (以下简称 PropBank) 的简单介绍。第三节是语料的预处理。第四到第六节是主要内容,其中包括特征选择,如何利用论元的上下文特

征提高分类准确率,以及如何设计贪心策略来选择特征窗口大小。第七节是有关的实验和结果、与前人研究结果的对比。第八节是展望。

2 中文 Proposition Bank 简介

中文 Proposition Bank (以下简称中文 PropBank) 是宾州大学仿照英文 PropBank 制作的中文语义角色标注语料库。它主要由两个资源构成: 1. **语义角色标注语料**。2. **动词框架**。其中资源 1 是 PropBank 的主要内容,具体标记了动词和其论元成分在中文 TreeBank 中的位置;资源 2 是一个支持性的内容,类似于词典,标记了所有出现在 PropBank 中的动词的子语类框架。

中文 PropBank 是在中文 TreeBank 的基础上添加了一个语义角色标注层,标记出来动词和对应论元在 TreeBank 中的位置。图 1 是 PropBank 中的一个例子(chtb_433.fid 第 1 句):

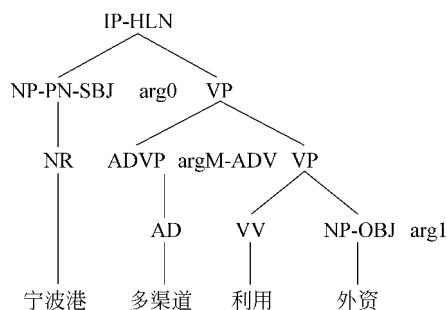


图 1 PropBank 的一个例子

在这个例子中,核心动词是“利用”。“利用”只有一个子语类框架,这个子语类框架包含两个论元成分:“利用(别人或物)的(人或物)”和“被利用的(人或物)”,分别对应原型施事和原型受事,在 PropBank 中标记为 arg0 和 arg1。在图 1 中,“宁波港”是“利用的”,“外资”是“被利用的”。除了原型施事和受事,在这个例子中还有一个论元成分“多渠道”,它是一个副词性的论元成分,表示“利用”的方式,标记为“argM-ADV”,其中“argM”是论元标记,“ADV”是一个二级的功能标记,二级标记在 PropBank 中还有很多,比如“MNR”(方式),“TMP”(时间)等。

中文 PropBank 的构建,其理论基础是连接理论。这个理论集中阐述了语义层面的角色如何实现为语法层面的句子成分,依据这一理论,很自然地会让人想到如何使用一种类似求逆的过程依据句子结

构得出“谓词—论元”结构。无论是语义角色的识别还是分类,都是在句法分析树上进行的,可以被看作是一个节点分类的问题。从某种程度上说,中文 PropBank 这个语料决定了我们的方法论。

3 语料的预处理

PropBank 是在 TreeBank 的基础上进行标注。Treebank 中的句法标注主要依据的是转换—生成理论,故而在标记中包含了大量由于移位(movement)而造成的空语类(null categories)。空语类的存在,使得移位的句法成分往往和其对应的空语类同时充当某个动词的同一个论元成分。此外,目前的句法分析器在句法分析结果中是不包含空语类的。这使得自动句法分析和人工标注的句法树结构不一致,从而使在其基础上生成的语义角色分类的结果不具备可比性。综合以上两个原因,在我们的工作中,空语类全部被删除了。

既然是语义角色分类,那么首先要解决的是有多少个类的问题。中文 PropBank 的理论基础是 Dowty 的语义角色原型理论,标记可以分为两级:一级标记是 arg0-4, argM, 共计 6 个。二级标记是形如 ADV, TMP 这样的功能性标记,共有 16 个。二级标记对 argM 是必需的,对于 arg0-4 是可选的。我们并没有直接使用 PropBank 中的论元标记,而是对其中的论元标记进行了化简,采用了表 1 中的分类。这样分类的依据是:首先,对于 arg0-arg4 这一类的论元,我们不妨称其为核心论元,虽然有些也包含二级的功能标记,但是毕竟是少数;并且功能标记不同的同类核心论元之间区别不大,所以二级功能标记对核心论元起的作用有限,我们就不再考虑细分了。但是 argM 这类论元,不妨称其为非核心论元,它的内部构成非常庞杂,一致性比较差,二级功能标记不同的 argM 差别很大。比如同样是 argM, argM-MNR 与 argM-TMP 的一致性就很低,所以对 argM 我们将其进行了细分。

表 1 中的标记,其中前 21 类都比较常见,arg0-4 是核心论元。包含施事、受事、与事等。argM 都是非核心论元,功能标记是进一步的细分。主要的有 EXT(程度),FRQ(频率),LOC(地点),MNR(方式),TMP(时间),TPC(话题)等^①。第 22 类在中文 PropBank 中没有任何说明,为了保证完整性,TBERR 这一类仍然保留了下来。

表 1 论元分类

1	ARG0	12	ARGM-DIR
2	ARG1	13	ARGM-DIS
3	ARG2	14	ARGM-EXT
4	ARG3	15	ARGM-FRQ
5	ARG4	16	ARGM-LOC
6	ARGM-ADV	17	ARGM-MNR
7	ARGM-ASP	18	ARGM-PRD
8	ARGM-BNF	19	ARGM-PRP
9	ARGM-CND	20	ARGM-TMP
10	ARGM-CRD	21	ARGM-TPC
11	ARGM-DGR	22	TBERR

4 特征选择

4.1 基准特征

参考了英文上面的实验和 Xue Nianwen 在中文上的研究^[16, 17],我们使用了以下特征模板作为我们的基准特征。它们可以分为简单特征和复合特征两类。简单特征又可以分为词特征、词性特征、句法特征。

简单特征

词特征

首词 该动词论元的短语中的第一个词

尾词 该动词论元的短语中的最后一个词

中心词 该动词论元的短语中的中心词,这里借鉴了 Collins^[18] 中的启发式规则,针对汉语的情况做出了一些调整

动词 与该论元成分对应的主要动词

词性特征

中心词词性 该论元成分的中心词的词性

短语类型 该论元成分的短语类型

句法特征

位置 该论元成分的短语在动词前还是动词后

路径 这个和 D. Gildea^[1] 等中提到的 path 一致。例如在图 1 中,“宁波港”的路径特征是“NR ↑ NP-PN-SBJ ↑ IP-HLN ↓ VP ↓ VP ↓ VV”。

语态 主动或者被动。本文采用了一种较为简

① 更详细的说明请参照中文 PropBank。

©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

单的判别方式, 即是否包含“把”、“被”

动词的框架 动词的父节点及其所有子节点构成的框架

围绕动词的 NP 从核心动词到最顶层节点之间各层节点周围围绕 NP (名词短语) 的情况, 比如图 1 中各论元成分的该特征都是(NP((V NP)))

路径中的 VP 数量 路径中包含的 VP (动词短语) 数量

到语态标记词的路径 该论元成分对应节点到语态标记词“把”、“被”的路径

右兄弟节点的短语类型 该论元成分对应节点的右兄弟节点的短语类型

复合特征

动词+短语类型

动词+中心词

短语类型+动词子语类框架简单表达式 动词子语类框架简单表达式与 Xue^[17] 中提到的 Frame 一致。如动词“保持”有两种子语类框架, 第一种包含三个核心论元, 是 arg0, arg1, arg2; 第二种包含两个论元, 是 arg0, arg1。那么其子语类框架简单表达式是: C3C2。

4.2 新特征

除了以上这些特征之外, 我们还用了一些与前人不同的特征, 有以下几种:

尾词的语义类 这里的语义类的信息是取自北京大学计算语言所的资源《现代汉语语义词典》¹⁹。如果尾词是名词, 则该特征的值是在《现代汉语语义词典》中的语义类, 否则为空。

中心词+动词子语类框架复杂表达式 动词子语类框架复杂表达式是将动词所有子语类框架都连结起来构成的表达式。例如动词“保持”的子语类框架复杂表达式是: CO_1CO_1_2^①。

短语类型+动词子语类框架复杂表达式

5 利用上下文特征

提取每个论元成分各自的特征进行学习标注, 这是语义角色标注的传统方法。在传统的方法中, 学习的单位是一个个单一的论元成分。同一动词的不同论元成分之间没有任何联系, 因此我们可以说, 这是一种单点分类的观点。

这样的处理方法好处是简单明了, 但是也存在一个很大的问题, 就是割裂了“单点”之间的相互关系。在汉语语义角色分类中, 具体体现为忽略了同一动词的不同论元之间的相关性。这种相关性在语言学上可以表现为以下一些规则:

- 1. 同一类型的论元在一个动词的论元结构中一般只能出现一次。
- 2. 论元的出现是有顺序的, 例如一般而言, 在非被动句中, 主语位置上是原型施事比是原型受事的可能性大得多。

语言学上的理论给我们以启示, 我们可以考虑通过利用同一个动词的不同论元之间的相关性来改善标记的结果, 英语上的实验也证实了利用论元相关性改善标记结果是可行的^[20]。比如同样以图 1 中的句子为例:

表 2 论元的相关性

宁波市	多渠道	利用	外资
arg0	argM-ADV	动词	arg1
特征集合 1	特征集合 2		特征集合 3

前人的做法是对各个论元成分, 使用各自的特征集合。比如以表 2 为例, 我们对“宁波市”使用特征集合 1, “外资”使用特征集合 3。而从论元相关的角度来看, “宁波市”、“多渠道”、“利用”、“外资”这四个节点构成了一个序列, 由于节点之间存在着相关性, 这样我们就可以利用相关论元的特征来改善标记结果。比如对于“宁波市”, 我们可以同时利用特征集合 1, 2, 3 的信息。我们可以这样理解: 特征集合 1 包含的信息从正面指明了“宁波市”可能是什么性质的论元, 而特征集合 2、3 所包含的信息则从反面揭示了“宁波市”不可能是什么性质的论元。因为同种性质的论元在同一个动词的论元结构中一般只出现一次。

为了利用上下文的信息, 我们引入了“窗口”的概念。窗口就是所利用到的相邻论元信息的范围, 下文我们一律使用 $[-m, n]$ 这样的式子来表示, 它的意义是: 将所有的论元成分按照在句子中的先后顺序线性排列, 对于其中第 i 个论元, 我们用来训练或者用于分类的特征集合是从第 $i-m$ 个到第 $i+n$ 个论元的特征集的并集。

① 子类框架复杂表达式的内部排列是有顺序的, 包含论元数少的框架在前, 多的在后。

6 特征窗口选择

通过对单个特征的研究,我们可以发现不同的特征在取得最高分类准确率时对应的窗口大小并不一样(详见 7.3.3),这给我们启示,就是是否对不同的特征取不同的窗口可以改善分类结果。

仍以图 1 中的句子为例:

表 3 特征窗口选择

宁波市	多渠道	利用	外资
arg0	argM-ADV	动词	arg1
首词 1	首词 2		首词 3
路径 1	路径 2		路径 3
.....

与之前对特征集设定窗口大小不同,在这里我们细化到具体的每个特征模板。比如对表 3 中的“多渠道”这个论元,“首词”这个模板我们只取首词 2 这个特征集合,“路径”则取路径 1、路径 2、路径 3 这三个特征集合。这样就做到了对不同类型的特征模板选择了不同的窗口大小。

我们使用了总共 20 个特征模板,如果遍历所有特征模板的窗口设置情况,计算的量过于庞大。为了简便起见,我们设计了一种贪心策略取寻找最佳的特征窗口设置。贪心策略的具体设计如下:

- 1 所有的特征窗口都设置为 $[-1, 1]$
- 2 For $i=0$ to 19
- 3 for 特征模板 i 的窗口设置从 $[0, 0]$ 到 $[-3, 3]$
- 4 if (结果变好) 保存特征模板 i 的这个窗口设置
- 5 End for
- 6 End for

我们希望通过这个贪心策略,选择一个局部最优的特征窗口设置,从而提高标注的准确率。

7 数据与实验

7.1 实验数据

本实验所用数据是宾州大学标注的中文 PropBank1.0 (LDC 编号为 LDC2005T23), PropBank1.0 标记了 Chinese TreeBank 5.1 的前 250k 语料,共标记了 4 865 个动词的 37 183 个论元结构。

文件编号从 chtb_001 到 chtb_931,中间有空缺。本文的试验就是在这个语料上进行的。实验使用如下方式划分语料:训练集语料 648 个文件(chtb_081 ~ chtb_899),测试集语料 72 个文件(chtb_001 ~ chtb_041, chtb_900 ~ chtb_931),开发集语料 40 个文件(chtb_041 ~ chtb_080)^①。这个语料的设置与 Xue (2008)相同,与 Xue(2005)不是完全相同。

7.2 最大熵分类器

本实验采用 Zhang Le 的最大熵分类器 MaxEnt,该分类器实现了包含高斯平滑的最大熵算法,采用 LBFGS 参数估计方法,可以很方便地处理多类划分的问题^②。

实验的参数设置如下:迭代次数 500,高斯平滑参数为 15。

7.3 实验结果

7.3.1 单点分类的结果

使用基准特征,在测试集上可以达到 94.08% 的准确率。加上了我们的新特征,准确率可以提高到 94.19%。

7.3.2 利用上下文特征的结果

如前所述,我们将从一个新的角度审视语义角色分类的任务,通过改变窗口大小改善分类的结果。表 4 显示的是分类准确率随窗口大小变化的情况。

表 4 不同窗口标记准确率(开发集上)

$[0, 0]$	$[0, 1]$	$[-1, 0]$	$[-1, 1]$	$[-1, 2]$
93.09%	93.36%	93.47%	93.93%	93.09%
$[-2, 1]$	$[-2, 2]$	$[-2, 3]$	$[-3, 2]$	$[-3, 3]$
93.00%	93.25%	93.59%	93.25%	93.05%

从表 4 中我们可以看到,在窗口大小为 $[-1, 1]$ 时语义角色分类取得最高的准确率,测试集上可以达到 94.69%,这相对于前人一般的处理策略,即窗口为 $[0, 0]$ 时的结果,有很大的提高。

7.3.3 单类特征标记准确率

对于单个特征模板提取的特征,我们也可以进行提取上下文特征的实验。对同一类特征采用不同的窗口,标注准确率可以有很大的变化。具体如

① 此外还去除了 一些 PropBank1.0 中与 TreeBank5.1 中对应不上的例子。

② 下载地址是: http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html

表 5 所示。

表 5 各类特征标记准确率(开发集上)

	[0, 0]	[0, 1]	[- 1, 0]	[- 1, 1]	[- 1, 2]	[- 2, 1]	[- 2, 2]	[- 2, 3]	[- 3, 2]	[- 3, 3]
1	26. 96%	29. 01%	30. 41%	30. 39%	30. 39%	32. 56%	31. 93%	31. 54%	31. 93%	31. 54%
2	30. 44%	36. 40%	34. 93%	38. 41%	39. 20%	38. 23%	38. 65%	38. 50%	38. 43%	38. 52%
3	33. 12%	47. 01%	39. 20%	50. 06%	50. 06%	50. 06%	52. 16%	51. 41%	52. 16%	51. 41%
4	34. 82%	47. 98%	40. 80%	50. 58%	50. 67%	50. 69%	52. 86%	52. 22%	52. 90%	52. 13%
5	36. 80%	50. 19%	41. 88%	53. 62%	55. 95%	55. 00%	56. 85%	57. 08%	57. 06%	57. 15%
6	37. 91%	52. 13%	43. 22%	53. 99%	55. 09%	54. 57%	55. 41%	55. 59%	54. 89%	55. 29%
7	45. 63%	52. 99%	47. 55%	53. 31%	56. 18%	53. 78%	55. 77%	56. 02%	56. 15%	56. 31%
8	50. 46%	61. 93%	51. 61%	63. 76%	64. 33%	63. 36%	63. 99%	63. 65%	63. 67%	63. 29%
9	59. 04%	61. 26%	70. 99%	71. 19%	71. 26%	71. 26%	71. 28%	71. 37%	71. 14%	71. 19%
10	60. 44%	60. 80%	63. 31%	64. 03%	62. 90%	64. 24%	63. 17%	63. 02%	62. 99%	62. 90%
11	60. 92%	69. 65%	65. 61%	70. 26%	70. 24%	70. 40%	70. 29%	70. 60%	70. 38%	70. 74%
12	65. 30%	68. 68%	70. 47%	72. 75%	72. 86%	73. 36%	73. 18%	72. 43%	73. 27%	72. 48%
13	69. 32%	71. 19%	72. 12%	72. 91%	72. 43%	73. 18%	73. 00%	72. 84%	72. 63%	72. 57%
14	69. 99%	70. 42%	72. 82%	73. 90%	74. 24%	73. 83%	73. 74%	73. 85%	73. 79%	73. 83%
15	70. 49%	71. 44%	72. 66%	73. 74%	73. 36%	74. 10%	73. 56%	73. 38%	73. 45%	73. 29%
16	84. 24%	84. 01%	84. 20%	84. 06%	83. 81%	84. 38%	83. 97%	84. 08%	84. 01%	84. 17%
17	85. 39%	86. 07%	85. 82%	86. 70%	86. 61%	87. 04%	86. 88%	86. 84%	86. 81%	86. 79%
18	85. 41%	85. 21%	86. 18%	85. 93%	86. 45%	85. 64%	86. 29%	86. 09%	86. 09%	85. 93%
19	87. 88%	88. 17%	88. 73%	89. 30%	89. 28%	89. 18%	89. 23%	89. 21%	89. 18%	89. 16%
20	87. 90%	88. 19%	88. 73%	89. 32%	89. 23%	89. 21%	89. 21%	89. 16%	89. 18%	89. 14%

表中各数字代表的特征是：
1 位置, 2 围绕动词的 NP, 3 语态, 4 到 语态标记词的路径, 5 动词的框架, 6 动词, 7 路径中的 VP 数量, 8 尾词的语义类, 9 中心词词性, 10 动词+中心词, 11 右兄弟节点的短语类型, 12 首词, 13 中心词, 14 中心词+动词子语类框架复杂表达式, 15 尾词, 16 动词+短语类型, 17 短语类型, 18 路径, 19 短语类型+动词子语类框架复杂表达式, 20 短语类型+动词子语类框架简单表达式。

通过上面的实验我们可以发现, 不同的特征取得标注最高准确率时对应的窗口大小是不同的, 有些特征在窗口较小时效果比较好, 有些则随着窗口增大不断提高。

7.3.4 选择特征窗口的结果
通过使用上述的贪心策略进行特征窗口的选择, 本系统最终可以将测试集上的标注准确率提高

到 95.00%。我们采用 χ^2 (自由度为 1) 显著性检验, 测试数据中论元总数为 7 714 个, 贪心策略进行窗口选择的结果比之前不考虑窗口设置做法(即窗口大小为[0, 0])的结果有显著提升($p < 0.05$)^①。

表 6 是各类论元的分类标记准确率。

表 6 各类论元标记准确率

	正例	总和	正确率
ARG0	1 995	2 046	97. 51%
ARG1	2 333	2 428	96. 09%
ARG2	202	260	77. 69%
ARG3	16	25	64. 00%
ARG4	3	5	60. 00%

① 依照 Pradhan 等(2005)的做法。

©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

续表			
	正例	总和	正确率
ARGM-ADV	1 440	1 490	96.64%
ARGM-ASP	1	1	100.00%
ARGM-BNF	20	23	86.96%
ARGM-CND	8	9	88.89%
ARGM-CRD	0	0	0.00%
ARGM-DGR	0	0	0.00%
ARGM-DIR	19	31	61.29%
ARGM-DIS	75	97	77.32%
ARGM-EXT	3	8	37.50%
ARGM-FRQ	0	0	0.00%
ARGM-LOC	286	308	92.86%
ARGM-MNR	230	248	92.74%
ARGM-PRD	0	0	0.00%
ARGM-PRP	44	46	95.65%
ARGM-TMP	648	675	96.00%
ARGM-TPC	5	12	41.67%
TBERR	0	2	0.00%
总计	7 328	7 714	95.00%

7.3.5 标注准确率综合比较

将前面的所有实验以及前人的实验结果综合起来,比较结果如表 7^①:

表 7 实验比较

基准特征	94.08%
所有特征	94.19%
所有特征+[-1, 1] 窗口	94.69%
所有特征+贪心策略选择窗口	95.00%
Xue(2005)	93.9%
Xue(2008)	94.1%

8 展望

语义角色标注是自然语言处理中一个较新的领域,汉语方面还有很大的研究空间。本研究只是语

义角色标注整个工作的一个组成部分,因为只涉及了其中的第二个阶段——语义角色分类,将来的工作重点将会放在语义角色标注的第一阶段——语义角色识别上面,使其与现有的工作结合起来,从而构建一个完整的汉语语义角色标注系统。我们希望通过汉语语义角色标注系统的建设,为自然语言处理的其他领域,比如对文本自动摘要,机器翻译等的研究提供帮助。

参考文献:

[1] D. Gildea, D. Jurafsky. Automatic labeling of semantic roles [J]. Computational Linguistics 2002, 28 (3): 245-288.

[2] F. C. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet project [C] // Proceedings of the 17th international conference on Computational linguistics. Montreal, Canada, 1998, 86-90.

[3] Paul Kingsbury and Martha Palmer. From TreeBank to PropBank [C] // Proceedings of the 3rd International Conference on Language Resources and Evaluation. Las Palmas, Spain, 2002, 1989-1993.

[4] Karin Kipper, Hoa Trang Dang, and Martha Palmer. Class-based construction of a verb lexicon [C] // Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence. Austin, Texas, USA, 2000, 691-696.

[5] Carreras X, Màrques L. Introduction to the conll-2004 shared task: Semantic role labeling [C] // Proceedings of the 8th Conference on Computational Natural Language Learning. Boston, MA, USA, 2004, 89-97.

[6] Carreras X, Màrques L. Introduction to the conll-2005 shared task: Semantic role labeling [C] // Proceedings of the 9th Conference on Computational Natural Language Learning. Ann Arbor, MI, USA, 2005, 152-164.

[7] S. Pradhan, K. Hacioglu, V. Krugler, W. Ward, J. H. Martin, D. Jurafsky. Support vector learning for semantic argument classification [J]. Machine Learning 2005, 60(1-3): 11-39.

[8] A. Moschiti. A Study on Convolution Kernels for Shallow Statistic Parsing [C] // Proceedings of the 42nd Meeting of the Association for Computational Linguistics. Barcelona, Spain, 2004, 335-342.

(下转第 68 页)

① 实验设置与 Xue(2005)有差异,因此数据不具备严格的可比性,但实验设置与 Xue(2008)完全相同。

Press, 2vols.

- [3] D. Stein and S. Wright. Subjectivity and Subjectivisation [M]. Cambridge: Cambridge University Press, 1995.
- [4] S.-M. Kim and E. Hovy. Determining the Sentiment of Opinions [C] // Proceedings of COLING-04 the Conference on Computational Linguistics (COLING-2004). Geneva, Switzerland; 2004: 1367-1373.
- [5] M. Gamon and A. Aue. Automatic Identification of Sentiment Vocabulary: Exploiting Low Association with Known Sentiment Terms [C] // Proceedings of the ACL- 2005 Workshop on Feature Engineering for Machine Learning in NLP. Michigan, USA; 2005: 57-64.
- [6] M. Hu and B. Liu. Mining Opinion Features in Customer Reviews [C] // Proceedings of Nineteenth National Conference on Artificial Intelligence (AAAI-2004). San Jose, USA; 2004.
- [7] 姚天昉, 聂青阳, 李建超, 李林琳, 陈柯, 付宇. 一个用于汉语汽车评论的意见挖掘系统[C] // 中文信息处理前沿进展—中国中文信息学会二十五周年学术会议论文集. 北京: 清华大学出版社, 2006: 260-281.
- [8] 邱立坤, 程薇, 龙志伟, 孙娇华. 面向 BBS 的话题挖掘初探[C] // 自然语言理解与大规模内容计算. 北京: 清华大学出版社, 2005: 401-407.
- [9] 潘国英. 论汉语动词重叠的主观性表达[J]. 修辞学习, 2008(1).
- [10] S.-M. Kim and E. Hovy. Extracting Opinions Opinion Holders and Topics Expressed in Online News Media Text [C] // Proceedings of the Workshop on Sentiment and Subjectivity in Text (COLING-ACL 2006 Workshop). Sydney, Australia; 2006: 1-8.
- [11] Y. Xia, K.-F. Wong, and W. Li. A Phonetic-Based Approach to Chinese Chat Text Normalization[C] // Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL 2006). Sydney, Australia; 2006: 993-1000.
-
- (上接第 26 页)
- [9] M. Zhang, W. Che, A. T. AW, C. L. Tan, G. Zhou, T. Liu, S. Li. A Grammar-driven Convolution Tree Kernel for Semantic Role Classification[C] // Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Prague, Czech Republic; 2007, 200-207.
- [10] 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.
- [11] 于江德, 樊孝忠, 庞文博, 余正涛. 基于条件随机场的语义角色标注[J]. 东南大学学报, 2007, 23(3): 361-364.
- [12] H. Sun, D. Jurafsky. Shallow Semantic Parsing of Chinese[C] // Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. Boston, USA; 2004: 249-256.
- [13] N. Xue, M. Palmer. Annotating the Propositions in the Penn Chinese Treebank [C] // Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing. Sapporo, Japan; 2003: 47-54.
- [14] 刘怀军, 车万翔, 刘挺. 中文语义角色标注的特征工程[J]. 中文信息学报, 2007, 21(1): 79-84.
- [15] 吕德新, 张桂, 蔡东风, 朱江涛. 沈阳航空工业学院学报[J], 2006, 23(1): 44-46.
- [16] N. Xue, M. Palmer. Automatic semantic role labeling for Chinese verbs [C] // Proceedings of 19th International Joint Conference on Artificial Intelligence. Edinburgh, Scotland; 2005, 1160-1165.
- [17] N. Xue. Semantic Role Labeling of Chinese Predicates [J]. Computational Linguistics, 2008, 34(2): 225-255.
- [18] M. Collins. Head-Driven Statistical Models for Natural Language Parsing [D]. Pennsylvania University, 1999.
- [19] 王惠, 詹卫东, 俞士汶. 现代汉语语义词典规格说明书[J]. 汉语语言与计算学报, 2003, 13(2): 159-176.
- [20] Z. P. Jiang, J. Li, H. T. Ng. Semantic Argument Classification Exploiting Argument Interdependence [C] // Proceedings of 19th International Joint Conference on Artificial Intelligence. Edinburgh, Scotland; 2005, 1067-1072.