

汉语框架语义角色的自动标注研究进展*

李济洪¹ 王瑞波¹ 王蔚林² 杨杏丽² 高亚慧² 李国臣³ 谷波³

1 山西大学 计算中心, 山西 太原 030006; 2 山西大学 数学科学学院, 山西 太原 030006

3 山西大学 计算机与信息技术学院, 山西 太原 030006

E-mail: lijh@sxu.edu.cn

摘要: 在给定句子中目标词及其所属框架的前提下, 本文以词为标注单位, 将语义角色标注形式化为词序列标注问题, 分别使用条件随机场和最大熵模型, 基于山西大学的 CFN 语料库, 研究汉语框架语义角色的自动标注。本文选取词、词性、位置以及它们的组合特征和窗口特征为候选特征集合, 并使用正交表来优选模型的特征模板。最终实验结果表明, 条件随机场模型优于最大熵模型。在测试集上, 条件随机场模型的 F 值达到 59.48%。若只识别语义角色的边界, 则 F 值为 70.40%。若给定语义角色边界, 语义角色分类的精确率为 80.4%。

关键词: 汉语框架网; 语义角色标注; 正交表; 特征选择

Research Progress of Chinese FN Semantic Role Labeling

LI Jihong¹, WANG Ruibo¹, WANG Weilin², Yang Xingli², GAO Yahui², LI Guochen³,
GO Bo³

1. Computer Center of Shanxi University, Taiyuan, Shanxi 030006;

2. School of Mathematics, Shanxi University, Taiyuan, Shanxi 030006;

3, School of Computer & Information Technology, Shanxi University, Taiyuan, Shanxi 030006

E-mail: lijh@sxu.edu.cn

Abstract: Given predicate word and its frame, this paper formalizes semantic role labeling of Chinese FrameNet to sequence labeling problem at word-level, and respectively employs Conditional Random Fields model and Maximum Entropy model to label semantic roles of Chinese FrameNet based on CFN semantic annotated Corpus. Furthermore, this paper constructs a candidate set of features, including word, part of speech, position etc, and then selects the best feature template using orthogonal array methodology. The experimental results show that, the performance of CRFs is better than ME model. On the test set, the F-measure of semantic roles labeling is 59.48% on the test data set. The F-measure of boundary identification subtask can achieve 70.40%. Based on human-corrected boundaries, the accuracy of semantic role classification subtask achieves 80.4%.

Keywords: Chinese FrameNet; semantic role labeling; orthogonal array; feature selection

1 前言

目前自然语言处理技术的瓶颈是语义的自动分析, 尤其是句义分析。句子中词汇的语义角色标注是对句子进行语义分析的一个重要步骤, 是句子命题义的形式化表示的一种实现方式。汉语框架语义角色的自动标注是基于汉语框架语义知识库(CFN)对中文句子实现句义的形式化表示的方法, 其深入的研究, 将为构建大规模语义语料库提供有效的工具, 将有力推动中文信息检索、自动问答、信息抽取等自然语言处理技术的发展。

自 Gildea 等人对英文 FrameNet 开展的自动语义角色标注的开创性研究^[6]以来, 国际上对于

*本文承国家自然科学基金(60873128)、国家 863 高技术研究发展计划资助项目(2006AA01Z142)和国家社会科学基金青年项目(07CYY022)的资助。

语义角色标注举行过多次评测,其中有 Senseval-3^[1]和 SemEval2007^[4],以及 CoNLL 会议主办的 SRL Shared Task 2004^[2]、2005^[3]、2008^[5]。其中,CoNLL SRL Shared Task 多数是以英文的 PropBank、NomBank 为语料,而 Senseval 系列主要使用英语 FrameNet 语料库为训练和测试语料。其中,Senseval-3 的语义角色标注任务,有 8 个队参加了评测,其中有 4 个队给出测试集上的测试结果,准确率 (precision) 在 80-90%之间,召回率 (recall) 在 65-78%之间,最高的达到 86.9%的准确率和 75.2%的召回率。

汉语框架网络 CFN (Chinese FrameNet)^[7]工程是以 Fillmore 的框架语义学^[8]为理论基础、以加州大学伯克利分校的 FrameNet 为参照、以汉语真实语料为依据的供计算机使用的汉语词汇语义知识库。从 2006 年至今,山西大学已构建框架 300 个,21600 条例句。

英语框架语义角色的标注主要是在完全句法分析之后,以句法成分为标注单元。目前汉语完全句法分析还不成熟,绝大多数分析器只对简单句有效,对多动词复句的分析尚不可行。根据框架元素标注的原则,框架元素的界限以语义理解为标准,只要其实际语义内容是目标词的支配成分,无论在句法上属不属于目标词的从属成分,都要划归为目标词的框架元素,因此,元素往往是复杂结构短语或子句。本文不计划按照英语框架语义角色标注的思路,而是将汉语框架语义角色自动标注看作是句子的以词为标注单元的序列标注任务,并分别使用最大熵模型和条件随机场模型进行 CFN 语义角色标注研究。文献[9][10]中对若干框架作了初步的自动标注实验,本文是在更大规模语料上的实验结果。

2 CFN 语义角色标注任务描述

本文的 CFN 语义角色标注的任务定为: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,标出该目标词所支配的语义角色 (CFN 中也称为框架元素,包括核心框架元素,非核心框架元素和通用非核心框架元素)。

本文将上述自动标注任务划分为语义角色边界识别 (识别出句子中语义角色块) 和语义角色分类 (对每个语义角色块的名称进行标注) 两步。本文使用双层 (边界标记+类型标记) IOB2 标记^[12]来标记语料库例句中每个词语,将任务转化为整个句子的以词为标注单元的序列标注问题。

3 特征选择及优化

根据语料库的现有标记状况,本文采用的候选特征包括:

词:指当前所标注的词条。

词性:指当前标注的词条的词性。

位置:指当前标注的词语与目标词之间的位置关系,包括“目标词左(L)”,“目标词右(R)”和“目标词本身(T)”三种位置关系。

目标词:指当前标注的词语所在句子的目标词元。

除上述四种特征外,本文还将这些特征的二元组合、三元组合以及相互组合特征及加入到候选特征集合中,具体见表 1 所示。另外,表 1 还给出了每种候选特征的 4 个可选窗口大小。

表 1 中,“-”表示不选择该特征,[-m, n] (m 和 n 为正整数) 表示窗口的大小,其中-m 代表当前标注单位的左边开大小为 m 的窗口, n 代表当前标注单位的右边开大小为 n 的窗口。

表 1 中特征的所有组合可以构成 $4^{10} \times 2$ 种特征模板,在所有模板上进行训练、测试,计算量很大,显然不现实。本文利用统计学试验设计的正交表选优法给出了模板选择的方法:将表 1 中的候选特征看作因子,窗口大小看作水平,把各种组合构成的特征模板看作全体“试验点”空间,这个问题转化为统计学的试验设计中的在全部试验空间中“均匀”选取试验点的优化问题,

可以使用正交表的均匀分散性来选点，其原理和方法在试验设计文献[13]中有详细的描述。本文选用正交表 $L_{32}(4^9 \times 2^4)$ 进行实验，并将“相邻词的位置”这个三元组合特征拆成三个 2 水平的特征(列)，三列的 2 水平分别对应不选取(-)和[-2, 0], [-1, 1], [0, 2](窗口大小)，因子排列分别对应表中的前 9 列和最后 4 列。

表 1 候选特征及其窗口大小

特征类型	窗口大小			
词	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
相邻词的二元组合	-	[-1, 1]	[-2, 2]	[-3, 3]
词性	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
词性的二元组合	-	[-1, 1]	[-2, 2]	[-3, 3]
词相对于目标词的位置	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
相邻词的位置的二元组合	-	[-1, 1]	[-2, 2]	[-3, 3]
词与其词性的搭配	-	[0, 0]	[-1, 1]	[-2, 2]
词与其位置的搭配	-	[0, 0]	[-1, 1]	[-2, 2]
词性与词的位置的搭配	-	[0, 0]	[-1, 1]	[-2, 2]
相邻词的位置的三元组合	-	[-2, 0]	[-1, 1]	[0, 2]
词与目标词的二元组合	-	[0, 0]		
目标词(必选特征)	[0, 0]			
一阶转移特征(必选特征)				

4 后处理方法

由于自动预测的 BIO 输出序列可能出现不符合 IOB2 的标记规范。本文以满足 IOB2 规范的概率最大的标注序列作为语义角色的自动标注输出结果。

输出序列要符合如下的约束：

(1) 输出序列必须符合 IOB2 的标记规范：第一个词的预测标记不能为“I”，序列中不允许出现“OI”等。

(2) 目标词所对应的输出标记必须为“O”。

(3) 输出序列第一层的 BIO 标记序列必须和给定的边界 BIO 标记序列完全一致。

语义角色边界识别实验要满足(1)(2)，角色分类步骤满足(1)(2)(3)。

5 实验结果及分析

本文从目前已有的 300 个框架中挑选出例句个数较多的 25 个框架的所有例句(分布见表 2)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。由于语料规模不大，为减少波动，本文任取 2 份作为训练集，其它 2 份作为测试集，共组成 6 份训练、测试集。最终的评价指标为 6 组实验，在各自的测试集上，所有语义角色块的准确率(precision)、召回率(recall)和 F-值的平均来评价标注系统的性能，同时还可以得到 6 组实验的标准误差。本文中，语义角色自动标注正确指角色的边界和角色名称均完全识别正确。

表 2 语料库所选的 25 个框架分布情况表

框架名	框架元素个数	例句个数	词元个数	框架名	框架元素个数	例句个数	词元个数
感受	6	569	61	因果	7	140	14
知觉特征	5	345	36	陈述	10	1603	87
思想	3	141	15	拥有	4	170	15
联想	5	185	19	适宜性	4	70	7
自主感知	14	499	27	发明	12	198	13
查看	9	320	31	计划	6	90	9

思考	8	283	33	代表	7	80	8
非自主感知	13	379	28	范畴化	11	125	10
获知	9	258	15	证明	9	101	11
相信	8	218	26	鲜明性	9	260	30
记忆	12	298	17	外观	10	106	13
包含	6	126	24	属于某类	8	74	7
宗教信仰	5	54	4	合计	200	6692	560

本文分别使用条件随机场模型和最大熵模型进行实验。对于两种模型均使用高斯先验约束，高斯方差参数统一设为 1.0。

5.1 语义角色边界识别的实验结果

对于汉语框架语义角色的边界识别，本文将 25 个框架的所有训练集进行统一训练，并在测试集上进行测试。表 3 中给出了六组实验的平均指标。表 3 中，两种模型的最优模板都为 0 模板，但是，条件随机场模型的 F 值达到了 70.40%，显著高于最大熵模型的 F 值。

表 3 在整个数据集上角色边界识别结果

条件随机场模型					最大熵模型				
模板	P	R	F	±sd(F)	模板	P	R	F	±sd(F)
0	74.42%	66.80%	70.40%	1.00%	0	71.01%	63.34%	66.96%	0.54%

表 4 给出了 0 号模板所对应的特征选取方式。

表 4 最优模板(0号)对应的特征选取方式

特征类型	窗口大小	特征类型	窗口大小
词	[0, 0]	词/位置	[-2, 2]
词的二元组合	[-2, 2]	词性/位置	[-2, 2]
词性	[0, 0]	目标词	[0, 0]
词性的二元组合	[-1, 1]	位置三元组	[0, 2]
位置	[-2, 2]	词/目标词	[0, 0]
位置的二元组合	-	目标词(必选)	[0, 0]
词/词性	[-2, 2]	一阶转移特征(必选)	

为了分析各特征的重要程度，本文使用正交表数据分析的极差分析方法，以平均的 F 值作为考察指标进行极差分析，结果如表 5 所示。

表 5 特征重要程度评价(条件随机场模型的极差分析)

特征描述	最大水平	最大均值	最小水平	最小均值	极差	特征描述	最大水平	最大均值	最小水平	最小均值	极差
位置三元组	2	0.6897	1	0.6558	0.0340	词/词性	3	0.6757	1	0.6703	0.0054
词性	1	0.6873	4	0.6534	0.0339	位置	2	0.6755	1	0.6708	0.0047
词	1	0.6781	4	0.6678	0.0103	词二元组	4	0.6750	1	0.6713	0.0037
位置二元组	2	0.6771	1	0.6677	0.0094	词性二元组	4	0.6745	2	0.6714	0.0031
词性/位置	1	0.6755	2	0.6667	0.0088	词/目标词	2	0.6731	1	0.6723	0.0008
词/位置	2	0.6758	3	0.6689	0.0069						

表 5 已经按照各类特征的极差值进行了降序排列，极差越大特征越重要。从表中可以看出，位置三元组特征的极差最高，表明位置三元组特征对结果影响最大；其次是词性和词，而这些特征为必选特征，表明窗口大小对他们影响大。

5.2 在给定角色块边界下, 语义角色分类的实验结果

当已知角色的边界进行语义角色分类实验时, 本文仍以词作为标注单位, 除了使用表 1 所列的特征外, 将每个词的所在角色的边界 BIO 标记, 作为特征加入到各模板中, 对每个框架分别训练, 以六组实验的测试集上的平均 F 值最大的来选择每个框架的最好模板。表 6 中给出了 25 个框架语义角色的总的分类精确率(accuracy)。

表 6 基于正确的边界的语义角色分类的实验结果

25 个框架	条件随机场模型		最大熵模型	
	Accuracy	±sd (Accuracy)	Accuracy	±sd (Accuracy)
总计	80.38%	0.38%	78.86%	2.25%

从表 6 中看出, 两种模型的总体精确率的置信度 95%的置信区间有交叉, 因此, 不能判定两种模型的精确率有显著差别。若分别考察不同类型框架元素, 结果如表 7:

表 7 给定边界下不同类型框架元素的分类结果

框架元素类型	条件随机场模型			
	P	R	F	±sd (F)
核心	83.11%	90.94%	86.85%	0.50%
非核心	71.52%	67.76%	69.58%	0.63%
通用非核心	74.80%	37.13%	49.60%	1.28%

5.3 先自动识别边界, 再进行语义角色分类的实验结果

若使用 5.1 的边界自动预测的结果, 再使用 5.2 的模型对角色进行分类, 结果见表 8。

表 8 先自动预测的语义角色的边界再分类的实验结果

25 个框架	条件随机场模型				最大熵模型			
	P	R	F	±sd (F)	P	R	F	±sd (F)
总计	62.87%	56.44%	59.48%	0.66%	59.55%	53.20%	56.13%	2.53%

可以看出条件随机场合最大熵两种模型的 F 值有显著差异。

表 9 先自动识别边界再分类的不同类型框架元素的结果

框架元素类型	条件随机场模型			
	P	R	F	±sd (F)
核心	64.75%	63.64%	64.19%	0.68%
非核心	56.62%	48.85%	52.45%	1.18%
通用非核心	59.85%	24.99%	35.26%	1.13%

6 结论与展望

本文基于山西大学的汉语框架语义知识库 (CFN), 分别使用条件随机场模型和最大熵模型进行了汉语框架语义角色的自动标注实验。模型以词为基本标注单元, 选择词、词性、位置、目标词, 及其相应的组合特征, 窗口特征, 采用统计学中正交表来筛选较优的特征模板。给定句子中的目标词, 以及目标词所属的框架情况下, 条件随机场模型的语义角色标注性能较高。在测试集上语义角色的标注准确率, 召回率, F1-值分别达到 62.87%(P),

56.44%(R), 59.48%(F1)。若只识别语义角色的边界,则 CRF 模型得到 74.42%(P), 66.80%(R), 70.40%(F1), 与最大熵模型有显著差异。若在已知角色的边界下,语义角色分类 CRF 模型得到 80.4%的精确率,与最大熵模型没有显著差异。

下一步,本文将在增加语料库中每个词元的例句个数,并在模型中加入组块特征,作深入研究,努力提高自动标注系统的性能。最终设计完成一个的框架语义自动标注器,为应用系统提供一个的语义分析工具。

致谢

本文的实验分析和撰写过程,一直得到刘开瑛老师的指导,在此表示感谢。

参 考 文 献

- [1] Litkowski K. Senseval-3 task: Automatic Labeling of Semantic Roles[C]. Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, 2004, 9-12.
- [2] Carreras X, Màrques L. Introduction to the CoNLL-2004 Shared Task: Semantic role labeling[C]. Proceedings of the CoNLL 2004, 2004, 89-97.
- [3] Carreras X, Màrques L. Introduction to the CoNLL-2005 Shared Task: Semantic role labeling[C]. Proceedings of the CoNLL 2005. 2005, 152-164.
- [4] Baker C, Ellsworth M, Erk K. SemEval-2007 Task 19:Frame semantic structure extraction[C]. Proceedings of the 4th International Workshop on Semantic Evaluation, 2007, 99-104
- [5] Surdeanu M, Johansson R, Meyers A, Marquez L, Nivre J. The CoNLL 2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies[C]. Proceedings of CoNLL-2008, 2008.
- [6] Gildea D, Jurafsky D. Automatic Labeling of Semantic Roles[J], Computational Linguistics, 2002, 28(3):245-288.
- [7] You L, Liu K. Building Chinese Framenet Database[C]. Conference on Natural Language Processing and Knowledge Engineering, 2005, 301-306.
- [8] Fillmore C J. Frame semantics and the nature of language[C]. In Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech, 1976, 280:20-32
- [9] 刘鸣洋, 由丽萍. 汉语感知词语的语义角色标注规则初探[M]. 内容计算的研究与应用前沿. 北京:清华大学出版社, 2007:320-325.
- [10] 刘开瑛.汉语框架语义网(CFN)构建现状[C].第四届全国学生计算语言学研讨会,2008.
- [11] Marquez L, Carreras X, Litkowski K C, Stevenson S. Semantic Role Labeling: An Introduction to the Special Issue[J]. Computational Linguistics, 2008, 34(2):145-159.
- [12] Ramshaw L A, Marcus M P. Text chunking using transformation-based learning[C]. Proceedings of the 3rd Workshop on Very Large Corpora, 1995, 88-94.
- [13] 中国现场统计研究会三次设计组.可计算性项目的三次设计[M].北京大学出版社,1985