

山 西 大 学
2009 届硕士学位论文

基于条件随机场模型的汉语框架语义 角色自动标注研究

#

作者姓名	王瑞波
指导教师	李国臣（教授）
学科专业	计算机软件与理论
研究方向	统计自然语言处理
培养单位	计算机与信息技术学院
学习年限	2006 年 9 月—2009 年 6 月

二〇〇九 年 六 月

Thesis for the 2009 Master's Degree of Shanxi University

Automatic Semantic Role Labeling of Chinese FrameNet Based On
Conditional Random Fields Model

Name	WANG Ruibo
Supervisor	Prof. LI Guochen
Major	Computer Software and Theory
Field of Research	Statistical Natural Language Processing
Department	School of Computer & Information Technology
Research Duration	September, 2006--June, 2009

June, 2009

目 录

第 1 章	引言	1
1.1	研究背景和意义	1
1.2	本文研究内容	2
1.3	国内外研究现状	2
1.3.1	FrameNet 相关语义角色评测任务介绍	2
1.3.2	文献综述	5
1.4	论文结构安排	7
第 2 章	汉语框架网络介绍及句法语义分析模型的构建	9
2.1	FrameNet 工程简介	9
2.2	汉语框架网络介绍	10
2.2.1	框架库	10
2.2.2	词元库	12
2.2.3	句子库	12
2.3	CFN 句法语义分析模型的构建	13
2.4	汉语框架语义角色标注任务描述	17
2.5	本章小结	17
第 3 章	条件随机场模型介绍	19
3.1	条件随机场模型	19
3.2	条件随机场模型的特征选取方法	20
3.3	条件随机场模型的参数估计方法	21
3.3.1	条件随机场模型权重参数的极大似然估计	21
3.3.2	条件随机场模型的极大似然估计缺陷	22
3.3.3	条件随机场模型参数的最大化后验估计	22
3.4	条件随机场模型的概率推理机制	23
3.4.1	归一化因子的计算公式	23
3.4.2	最优标注序列的寻找	25
3.5	IOB2 标注策略及不合理序列的处理	26
3.6	条件随机场模型的改进	27
3.7	条件随机场工具包介绍	28
3.8	本章小结	29
第 4 章	汉语框架语义角色标注的基线模型	31
4.1	基于词的汉语框架语义角色标注模型	31
4.1.1	模型的形式化描述	31
4.1.2	模型标注范围的选取	31
4.2	特征描述及选取方法	34
4.3	语料选取	37

4.4 语料切分	38
4.5 评价指标	39
4.6 特征重要性分析	39
4.7 实验结果	40
4.7.1 “陈述”框架动词词元的实验结果	40
4.7.2 “陈述”框架名词词元的实验结果	42
4.7.3 “陈述”框架的语义角色标注实验结果	44
4.8 实验分析	44
4.8.1 特征的极差分析	44
4.8.2 模型的性能分析	45
4.8.3 “陈述”框架语义角色标注性能曲线	47
4.9 其它框架的实验结果	48
4.10 本章小结	50
第5章 融合汉语基本块信息的汉语框架语义角色标注模型	51
5.1 汉语基本块描述体系及基本块自动标注性能	51
5.2 基本块和汉语框架语义角色的关系	52
5.3 融合基本块信息的汉语框架语义角色标注模型	54
5.4 基于词的模型	55
5.4.1 模型的特征描述及选取	55
5.4.2 模型的实验结果	57
5.4.3 实验分析	63
5.4.4 其它框架的实验结果	65
5.5 基于基本块的模型	66
5.5.1 模型预处理	67
5.5.2 模型的特征描述及选取	67
5.5.3 模型的实验结果	68
5.5.4 实验分析	73
5.5.5 其它框架的实验结果	75
5.6 模型的比较与融合	76
5.7 本章小结	78
第6章 总结及展望	81
6.1 论文总结	81
6.2 下一步研究计划	82
参考文献	85
附录 A: 山大 973 词性标注体系及其与清华大学词性映射关系	91
附录 B: 实验用正交表及 PB 表	93
发表文章目录	97
致 谢	99
个人简历	101

CONTENTS

CHAPTER 1 INTRODUCTION	1
1.1 Research Background and Significance	1
1.2 Works of Thesis	2
1.3 Status Quo at Home and Abroad	2
1.3.1 Evaluation Workshops Involved FrameNet.....	2
1.3.2 Review of Research Works.....	5
1.4 Structure of Thesis.....	7
CHAPTER 2 INTRODUCTION TO CFN AND ARCHITECTURE OF PARSING MODEL	9
2.1 Introduction to FrameNet Project.....	9
2.2 Introduction to Chinese FrameNet	10
2.2.1 Frames DataBase	10
2.2.2 Lexical Units DataBase	12
2.2.3 Annotated Sentences Corpus	12
2.3 Architecture of CFN Sytax-Semanteme-Integrated Model	13
2.4 Description of CFN Semantic Role Labeling Task	17
2.5 Conclusion.....	17
CHAPTER 3 INTRODUCTION TO CONDITIONAL RANDOM FIELDS MODEL.....	19
3.1 Conditional Random Fields Model	19
3.2 Feature Selection Methods of CRF	20
3.3 Parameters Estimation of CRF	21
3.3.1 Maximum Likelihood Estimation of Parameters	21
3.3.2 Disadvantages of MLE	22
3.3.3 Maximum A posteriori Estimation of Parameters	22
3.4 Probabilistic Inference of CRF	23
3.4.1 Computation of Normalization Factor.....	23
3.4.2 Search of Best Labeling Sequence	25
3.5 IOB2 Representation and Postprocess of Irregular Output	26
3.6 Improvement of CRF Model	27
3.7 Introduction to CRFs Toolkits	28
3.8 Conclusion.....	29
CHAPTER 4 BASELINE MODEL OF CFN SEMANTIC ROLE LABELING	31
4.1 Word-Based Chinese FrameNet SRL Model.....	31
4.1.1 Formalized Representation of Model	31
4.1.2 Selection of Labeling Scope of Model	31
4.2 Feature representation and Selection.....	34
4.3 Selection of Experimental Corpus.....	37
4.4 Split of Experimental DataSet.....	38

4.5 Evaluation Metrics	39
4.6 Importance Analysis of Features	39
4.7 Experimental Results.....	40
4.7.1 Result of Model of Verbal Target Words In Statement Frame	40
4.7.2 Result of Model of Nominalized Target Words In Statement Frame	42
4.7.3 Result of Statement Frame	44
4.8 Analysis of Experiments.....	44
4.8.1 Range Analysis of Features	44
4.8.2 Performance Analysis of Features	45
4.8.3 Performance Curves of Statement Frame.....	47
4.9 Results of Other Frames	48
4.10 Conclusion.....	50
CHAPTER 5 BASECHUNK-INTEGRATED CFN SRL MODEL	51
5.1 Chinese Base Chunk Scheme and Parser's Performance.....	51
5.2 Relationship between BaseChunks and Frame Elements.....	52
5.3 BaseChunk-Integrated Semantic Role Labeling Models	54
5.4 SRL Model Based on Word.....	55
5.4.1 Representation and Selection of Features.....	55
5.4.2 Results of Model.....	57
5.4.3 Analysis of Results	63
5.4.4 Results of Other Frames	65
5.5 SRL Model Based on BaseChunk	66
5.5.1 Preprocess of Model.....	67
5.5.2 Representation and Selection of Features.....	67
5.5.3 Results of Model.....	68
5.5.4 Analysis of Results	73
5.5.5 Results of Other Frames	75
5.6 Comparison and Integration of Models	76
5.7 Conclusion.....	78
CHAPTER 6 CONCLUSIONS AND FURTHER DIRECTIONS.....	81
6.1 Conclusions of Thesis.....	81
6.2 Further Directions.....	82
REFERENCES.....	85
APPENDIX A.....	91
APPENDIX B.....	93
PUBLICATIONS.....	97
ACKNOWLEDGEMENT	99
RESUME	101

摘 要

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。

考虑到目前汉语完全句法分析还不成熟,按照英语以句法成分作为标注单元的做法将难以实现。因此,本文将汉语框架语义角色自动标注看作是以词为标注单元和以基本块为标注单元的整个句子的序列标注任务,并分别使用条件随机场模型进行实验。本文的语义角色标注任务为:给定句子中的目标词及其所属框架,自动标出相应的框架语义角色。

本文从目前 CFN 的 200 多个框架中选取例句数相对较多的 25 个框架的所有例句按 5:5 构成训练、测试集。

对于基于词的标注模型,本文构造了词层面的特征(当前词、当前词的词性、当前词的位置及其组合),利用正交表来选择最优的特征模板。在测试集上,基于词的标注模型的 F 值达到了 62.50%。进一步,本文使用清华大学的基本块分析器对所有例句进行自动分析,将基本块特征加入到 baseline 模型中,测试集上 F 值有 1.15% 的上升,达到 63.65%。对结果的错误分析表明:词特征稀疏可能是导致性能不高的主要原因。

对于基于基本块的标注模型,本文以自动的基本块标注信息中构造若干浅层句法特征,同样利用正交表优选特征模板。在测试集上,最优模板的 F 值达到 62.11%。较 baseline 模型有约 0.4% 的下降,测试集的 25 个框架中多个框架的 F 值,较基于词的标注模型都有显著下降。主要的原因是基本块的标注性能较低。

最后,本文从语义角色标注模型的角度和机器学习的角度对研究过程中所遇到的一些问题进行了分析和归纳,指明进一步的研究方向。

关键词: 语义角色标注; 汉语框架网络; 条件随机场模型; 汉语基本块

ABSTRACT

Currently, the bottle-neck of natural language processing technologies is the automatic semantic parsing, especially the parsing of sentential meaning. As an important step of semantic analysis, semantic role labeling of lexical units is one of the formalized representations of propositional meaning of the sentence. Automatic semantic role labeling based on Chinese FrameNet is a formalized representation of Chinese sentences. The automatic construction of large scale semantic corpus and the improvement of natural language processing technologies, e.g. Chinese information retrieval task, Chinese question answering system, Chinese information extraction technique and etc., will be achieved along with the further research of automatic semantic role labeling based on Chinese FrameNet.

Due to the lack of an accurate Chinese syntactic parser, it is infeasible to adopt the SRL approach on English FrameNet. Therefore, we formulate the Chinese FrameNet semantic role labeling problem as a sequence labeling task, and use conditional random fields model to solve the problem at the word-level and base-chunk-level respectively. In this thesis, the task of Chinese FrameNet semantic role labeling is defined as: Given a Chinese sentence, a target word and its frame, identify the boundaries of frame elements within the sentence and label them with an appropriate frame element name.

We select 25 Frames from current Chinese FrameNet Corpus in terms of the count of sentences, and split them into training set and testing set according to the proportion of 5:5.

In the word-based labeling model, we construct several features at word-level, and employ orthogonal array to select the best feature template. The model achieves 62.50% F-measure, which is the baseline result we get. Furthermore, we parse all sentences in the corpus using the TsingHua Base-Chunk parser, and employ the base-chunk features into word-based

labeling mode. The result on the test data set achieves 63.65% F-measure, which increase around 1% to the baseline result. The experimental analysis demos that the main effect of the model performance is the sparsity of word feature.

In the base-chunk-based labeling model, several shallow syntactic features are constructed based on the automatic labeling information of Chinese Base-Chunk. Similar to the word-based labeling model, we employ orthogonal array to select the best feature template. On the test data set, the best-chunk-based labeling model achieves 62.11% F-measure, which is decreased by 0.4% to the baseline result. The F-measures of the majority of 25 Frames have significant decreases compared with the baseline results. The main reason is the accuracy of Chinese Base-Chunk parser is relatively low.

Finally, we conclude the thesis and summarize some problems in the perspective of the semantic role labeling model and machine learning technology, and provide some directions for further research.

Key words: semantic role labeling; Chinese FrameNet; Conditional Random Fields Model; Chinese Base Chunk

第1章 引言

1.1 研究背景和意义

随着词法分析技术和句法分析技术的日趋成熟，自然语言处理的重点逐步转变到语义分析任务上来。语义分析指的是根据句子的句法结构和句中每个实词的词义推导出能够反映这个句子意义（即句义）的某种形式化表示。语义分析的研究可以为语用分析和一些应用型的任务奠定基础。

目前，语义分析仍然停留在词义的形式化描述和分析阶段。词义分析的主要目的是针对句子中的谓词，形式化地描述出句子中由该谓词所刻画的事件和场景的一些基本要素，如，参与者、地点、时间等。其中，“参与者”，“地点”，“时间”等被称为语义角色。这种通过语义角色的方式来分析谓词词义的方法被称为浅层语义角色标注。它的基本目标^[1]是根据预先给定的谓词的语义角色列表，准确地分析一个谓词和与它有关的要素和属性之间的语义关系。

随着自然语言处理中统计机器学习方法的兴起，很多大规模具有语义信息的语料资源被建立，这为使用机器学习方法进行语义角色标注成为可能。目前，英语方面的语义资源较为丰富和成熟，其中一些已经具有完善的体系和规模较大的标注实例。例如，FrameNet^[2]，PropBank^[3]和 NomBank^[4]。

以 Fillmore 先生提出的框架语义理论为基础，伯克利分校开发了 FrameNet，并针对 BNC 语料库进行语义角色标注。从 2002 年 6 月发布自今，框架网络数据库中一共标注了 1,462 个目标谓词(927 个动词，339 个名词和 175 个形容词)。例句个数达到 49,000 句。与 FrameNet 不同的是，PropBank 基于 Levin 的语义理论体系，针对英文句子中的核心动词进行标注。并且使用框架文件(framfile)来刻画每个词元所具有的义项。对于多义动词，框架文件中使用不同的框架集(frameset)对该词语的每个义项进行区分。通过大量的标注实例发现，PennBank 中每个动词的核心语义角色不超过 6 个，它们被标记为 Arg0-5 六种，其中，Arg0 通常表示动作的施事，Arg1 通常表示动作的影响等，Arg2-5 根据谓语动词不同会有不同的含义。

PropBank 仅将动词作为谓词来标注它的谓词论元结构，而没有考虑名词作为目标词时所具有的语义角色。针对这个缺陷，纽约大学的研究人员开发了 NomBank 。与 PropBank 不同的是，NomBank 将 Penn 树库中的名词作为谓词来标注它所对应的语义角色。

汉语方面比较知名的语义角色标注库是汉语框架语义知识库^[5], 汉语 PropBank^[6]以及汉语 Nombank^[7]。

汉语框架网络 (CFN, Chinese FrameNet) 工程是以 C.J.Fillmore 的框架语义学为理论基础、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。汉语框架网络知识库包括框架库、句子库和词元库三部分组成。目前, CFN 已经对 2610 个词元构建了 230 个框架, 标注了 15000 条句子, 包含认知领域用词、科普文章常用谓词以及部分中国法律用词。

宾州大学开发的汉语 PropBank 主要仿造了英文 PropBank 的标注体系, 并针对汉语 Penn TreeBank 中的动词进行了相应的语义角色标注。汉语 Nombank 也是宾州大学研制的, 类似于汉语 PropBank, 它将传统英语 Nombank 的标注框架移植到了对汉语名词性谓词的标注上。

语义角色标注 (Semantic Role Labeling, SRL) 是目前语义分析的一种主要实现方式。语义角色标注是对自然语言理解的一个有益的探索, 是信息抽取、机器翻译、问答系统等多种自然语言处理技术的重要基础, 而且对以后进行深层语义分析以及篇章理解的研究有一定的意义。文献[8]将 FrameNet 的语义信息融合到了问答系统之中以提高问答系统的精度。文献[9]将汉语框架语义信息引入到了汉语旅游问答系统之中, 为提高系统的性能奠定了语义基础。文献[10]将汉语框架语义信息作为特征加入到了中文问题分类任务中, 使得问题分类精度有所提升。这些研究工作表明了语义信息的融合对问答系统等应用型系统性能的提升有着积极的作用。诸如此类的工作还有很多, 这里不再一一引出。

1.2 本文研究内容

本文主要研究汉语框架语义角色标注问题, 重点放在与框架有关的语义角色的自动标注问题的研究上。文中首次采用了山西大学开发的汉语框架网络知识库 (Chinese FrameNet, 简称 CFN) 作为本文的语料资源。汉语框架语义信息对于进行信息抽取、信息检索、中文阅读理解问答系统等多种自然语言处理任务的研究具有较大的实用意义。

本文采用统计机器学习的方法进行汉语框架语义角色自动标注。其中, 本文主要采用了条件随机场模型^[11]自动预测汉语句子中的框架语义角色信息。条件随机场继承了最大熵模型的优点, 可以有效的综合各种具有依存关系的特征信息, 根据训练集的不同自动调整模型的权重参数, 并对未见过的新的测试样例做出准确的预测。

本文首先将框架语义角色标注建立在词层面，并将框架语义角色自动标注任务看作是对一个词语序列中的每个词选择正确的语义角色标记的过程。在此基础上，本文将词层面的信息转化成相应的特征融合到条件随机场模型中。最终，本文得到了一个初步的汉语框架语义角色标注模型。在 25 个框架上的实验结果表明，汉语框架语义角色标注的总体 F 值可以达到 60% 以上。

基本块信息的融合可以为语义角色的自动预测提供更为丰富的信息。为此，本文将自动标注的基本块信息加入到汉语框架语义角色标注模型中。通过分别使用词语和基本块作为标注单位，本文分别建立了“基于词的标注模型”和“基于基本块的标注模型”。但是，基本块信息的加入并不能显著提高框架语义角色标注模型的性能。

除了上述两方面主要的工作外，本文还初步进行了以下几方面的工作：

第一，提出了汉语框架句法语义分析模型的完整的架构体系。

第二，使用正交表的方法进行了大量的特征选择实验。这为进一步研究“基于正交表的特征选择方法”提供了大量供分析使用的实验结果。

第三，通过实验结果及相应的理论分析，总结出条件随机场模型的一些缺陷。这为进一步改进条件随机场模型提供了参考。

1.3 国内外研究现状

目前，自动语义角色标注任务已经引起了国内外众多学者的广泛关注。许多人尝试使用机器学习的方法来进行研究。自 2002 年开始，国际上已经针对语义角色标注任务进行了多次评测，并且在《Computational Linguistic》和《Machine Learning》等核心期刊上也收录了大量的相关文献。特别值得关注的是，2008 年自然语言研究的国际权威学术期刊《Computational Linguistic》出版了一期关于语义角色自动标注研究的专辑。国内近年来的重要学术会议上，语义角色标注的研究也越来越得到同行的重视，已成为研究热点。在大规模语义知识库的构建、机器翻译、信息提取、问答系统、信息检索等应用领域语义角色的自动标注技术都有着广泛的应用，其深入的研究对自然语言处理技术的整体发展意义重大。

为了充分分析目前语义角色标注任务的趋势，本文对国际语义评测活动中涉及到 FrameNet 评测的相关任务进行详尽的阐述。在此基础上，本文对语义角色标注研究中一些重要的研究工作及相关文献进行了对比和分析。

1.3.1 FrameNet 相关语义角色评测任务介绍

2004 年至今,国际上已经针对语义角色标注开展过多次评测。其中,CoNLL 会议主办的 SRL Share Task 2004^[12]、2005^[13]以及 CoNLL-2008^[25],主要是以 PropBank 和 NomBank 为评测语料。而 Senseval 主办的 Senseval-3^[14],SemEval-2007^[15]以及即将召开的 Semeval-2010^[16]主要是以 FrameNet 为训练和测试语料。为了清晰地把握英文 FrameNet 的发展方向和评测重点,本文主要针对 Senseval 会议中涉及语义角色标注的相关任务进行详细地阐述。

1. Senseval-3 评测任务:语义角色自动标注 (Automatic Labeling of Semantic Roles)

Senseval-3 中关于语义角色自动标注任务的定义是:给定一个句子,句子中的目标词和它所属的框架,自动的识别出句子中所有的框架元素并且标注出各个框架元素的类型。

这个任务主要是对 Gildea 和 Jurafsky 的基础工作^[27]的扩展。任务所使用的数据均来自于 FrameNet 的手工标注数据集。评价系统使用的指标主要是框架元素识别的准确率、召回率以及自动识别出的框架元素与正确框架元素位置的重叠程度。实验数据来自于从 FrameNet 数据库中随机抽取出来的 40 个框架的所有标注例句,共 8002 句。

2. SemEval-2007 评测任务:框架语义结构抽取 (Frame Semantic Structure Extraction)

较 Senseval-3 来说,Semeval-2007 的任务的难度骤然增加。框架语义结构抽取任务主要有以下四个子任务组成:

- 1) 框架识别:识别出句子中能够激起框架的所有词语和短语,并且给它们标注上正确的框架信息(类似于 WSD 任务)。
- 2) 最近框架检测:对在测试集中出现,但未在训练集中出现的框架,自动检测出与该框架最近的已知框架。
- 3) 语义角色标注:识别出句子中目标词元所具有的框架元素的边界和类型。
- 4) 核心依存图抽取:根据句子的框架语义角色标注结果,自动构建出该句子的核心依存图。

该任务使用 FrameNet 的所有词典编撰式的标注例句和全文标注例句作为训练集,测试集使用了 3 篇新的全文标注。评价体系中采用了针对语义角色标注的标记评价,以及针对核心依存图的依存图评价。由于测试集中存在未知框架,因此在评

价体系中还引入了部分打分策略。这里不再展开叙述，感兴趣的读者可以参阅[15]。

3. SemEval- 2010 评测任务：话题的事件及其参与者链接（Linking Events and their Participants in Discourse）

SemEval-3 的语义角色标注任务和 Semeval-2007 的框架语义结构抽取任务将语义信息的表达看作是一个句子内部的任务。然而，在围绕同一个话题的句群中，不同句子的论元以及结构之间是相互影响的。SemEval-2010 中，话题的时间及其参与者链接任务的主要目的是将一段表述同一话题的文本中不同句子之间的语义论元和结构链接到一起。特别是要找到在句子内部被标为缺省，却在上下文环境中出现的那些语义角色，并将它们链接到一起。

该评测任务在给定目标词和相应框架的基础上，主要分为如下三个子任务：

- 1) 语义角色识别：找到句子中目标词所支配的语义角色。
- 2) 语义角色标注：为目标词所支配的语义角色标注上正确的角色名称。
- 3) 缺省成分链接(NI linking)：在上下文中将缺省成分链接到一起。

参与者也可以只进行第 3 个子任务的评测，此时，正确的语义角色标注信息将会被提供。评测数据主要来自小说领域的一些标注文本。训练集将会被标注上正确的语义论元结构以及缺省成分的链接信息。需要说明的是，语义论元结构的标注主要采用了 FrameNet 和 PropBank 两种体系，参与者可以根据自己喜好任选其一。

从上述三个评测任务的对比中可以看出，语义分析目前虽然仅仅停留在对单个词语的分析，但是，评测任务越来越注重使用句子中各个词语的语义关系来勾勒出一个句子的部分语义信息（核心依存图），甚至是一个篇章中的部分语义信息（缺省成分链接）。

1.3.2 文献综述

英语的语义角色自动标注已经有多年的研究，其中，绝大多数是基于英文的 PropBank、NomBank 语料。这些研究主要有 J.Chen^[17], Thompson^[18], Robert^[19], Kadri Hacioglu^[20], Pradhan^[21], M. Surdeanu^[22], Sameer S. Pradhan^[23]等人的工作。刘挺等^[24]基于英文 PropBank，以句法成分为基本标注单元，使用最大熵分类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。CoNLL-2008 Shared Task^[25]除要求角色标注之外，增加了谓词的识别和词义排歧，以及论元中心词与谓词的语法依赖关系（syntactic dependency）和语义依赖关系（semantic dependency）的分析，最好的评测结果 F1-值达到 84.86%。国内有 5 支队伍参加了评测，Wanxiang Che 等^[26]的评测结果最好达到 82.66%的 F1-值，在所有 19

支参加评测的队伍中排名第二。

基于英文 FrameNet 的语义角色标注的研究最早是 Gildea 等人^[27]的工作。他们针对每一个词元建立相应模型，并在句子完全句法分析树上，选取了大量的特征，其中包含短语类型、句法功能、位置、语态、中心词、路径等，最终使用条件概率估计的方法，在 FrameNet 测试集上进行了实验，取得了 65% 的准确率(precision)和 61% 的召回率(recall)。Senseval-3 的 SRL Task 的主要内容是语义角色标注，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。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识别（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。有 3 个队提交了评测结果，在 Task-19 给定的测试集上，框架的识别的最好结果为：62.61%（召回率），77.31%（准确率），69.18%（F1-值）；框架识别再加上相应语义角色标注的最好结果为：41.52%（召回率），61.06%（准确率），49.43%（F1-值）。

汉语的语义角色自动标注的较早研究是 Xue(2005)^[28]基于中文 PropBank 的自动标注研究工作。Nianwen Xue^[29]基于中文 PropBank 深入研究了语义角色的自动标注问题，在语料库中已手工标记好的句法分析基础上，得到了 94.1% 的 F1-值。但如果采用自动的句法分析，却只有 71.9% 的 F1-值，结果显著地低于相应英文的结果。PropBank 上的语义角色标注任务主要是针对每个句子的核心动词进行标注，语义角色的定义总共有 50 多个，其中主要角色只有 6 个，且所有动词的主要角色有相同的标记 Arg0-Arg5。因此，相对于框架语义角色标注任务来说，PropBank 的语义角色标注任务有较为丰富的资源，语义角色类型少，所有目标词的标记相同有利于自动标注模型的训练。Nianwen Xue 的实验也揭示出，Arg0，Arg1 的自动标记的结果要好于其他角色的标注，一个可能的原因是 Arg0，Arg1 角色的涵义具有同质性，如 Arg0 绝大多数情况下是指 Agent，利于统计模型的训练；而其他的角色虽然标记相同，其涵义却差别很大。正如袁毓林^[30]指出：“中文 PropBank 中，论元标记 Arg2 至 Arg5 对于语义角色的严重超载，使用这种标注语料来训练角色自动标注系统，其性能势必会受到影响”。事实上，PropBank 中语义角色类型忽略了语言表达中的细节，词汇义项的描述显得粗糙，其使用价值受到怀疑。

1.4 论文结构安排

本文的组织结构如下：

第2章首先介绍了汉语框架网(CFN)的理论框架以及目前的构建情况。基于CFN句子标注规范，本文提出了相应的CFN句法语义分析模型的构建设想，并给出了一个模型的概貌以及各个模块的详细介绍。最后，提出了本文所研究的汉语框架语义角色标注的定义及它在CFN句法语义分析模型中所占的位置。

第3章着重介绍了条件随机场模型的理论框架，以及在实现时所使用的参数估计方法。虽然，条件随机场模型被广泛使用到了序列标注和分割任务中，但是它也具有很多缺陷，例如，时间复杂度高等。本文在本章的最后给出了目前已有的一些对条件随机场模型的改进方法。

第4章使用词语作为标注单位初步构建了一个汉语框架语义角色标注基线模型。在这个基线模型中，本文仅使用了词语层面的信息，并将它形式化成不同的特征。为了客观地评价基线系统，本文充分分析了每个特征所起的作用，并给出了目前汉语框架语义角色标注的一个初步结果。

第5章进一步将汉语基本块信息引入到汉语框架语义角色标注模型中。本章首先介绍了汉语基本块的描述体系以及汉语自动基本块分析器的性能，并分析了汉语基本块和汉语框架语义角色之间的关系。以此为基础，提出了标注单位不同的两种汉语框架语义角色标注模型，即，“基于词的标注模型”和“基于基本块的标注模型”。随后，大量的实验结果以及相应的实验分析围绕着两个语义角色标注模型给出。实验结果表明，基本块信息的融合并不能显著地提高汉语框架语义角色标注的性能。

第6章针对汉语框架语义角色自动标注所作的工作进行了总结，并从语义角色标注的角度和机器学习的角度对研究过程中所遇到的一些问题进行了分析和归纳，以便于下一步工作的开展。

第2章 汉语框架网络介绍及句法语义分析模型的构建

随着自然语言处理研究的深入,语义关系的研究在自然语义处理领域越来越受到重视。目前,语义分析主要借助语义角色来刻画句子中谓词与它相应的参与者和属性之间的语义关系。自从 Fillmore 于 1968 年作出的奠基性工作之后,很多语言学家把目光转到了语义角色的相关研究上来。虽然在诸如施事(Agent)、受事(Patient)等一些重要的语义角色上,语言学家们可以达成一致的意見,但是却没有谁能给出一个完整的语义角色列表。其中,不同的语言学家们给出的语义角色的列表格式各样。语义角色列表的一个极端是包含了大量的与领域相关的语义角色。例如,在订机票系统中可以使用到“出发地(ORIG_CITY)”、“目的地(DEST_CITY)”和“出发时间(DEPART_TIME)”等语义角色。而另一个极端却仅提出了两种语义角色,即:“元施事(PROTO-AGENT)”和“元受事(PROTO-PATIENT)”。当然,处于两端之间也有许多理论,如二十世纪六七十年代,由 C.J.Fillmore 提出的格语法便是其中的一种。格语法最初仅使用数个格角色来表示语句中不同语义元素。但是,由于没有统一的划分和归类标准,格的数量不断增多,这样用格角色来表达语义也变得复杂。到了二十世纪末 Fillmore 提出了框架语义学,该理论认为谓词的句法和语义组合的可能性应该通过“框架”来描述,继而在框架语义学的基础上提出了框架网络描述体系。

2.1 FrameNet 工程简介

框架网络(FrameNet, 简称 FN)是伯克利分校基于 Fillmore 先生的框架语义学理论所构建的基于语料库的计算词典编纂工程。它的主要目的是研究英语中句法功能和语义结构之间的关联关系,并为句子和篇章语义信息的部分理解奠定基础。目前,框架网络数据库中共标注了现共标注了 1,462 个目标谓词(927 个动词,339 个名词和 175 个形容词)。例句个数达到 49,000 句。

框架语义学从认知语言学的观点出发,对句子中的词语的语义信息进行了分析。它认为,句子中的每一个谓词都可以激起一个框架,而该谓词的含义可以通过它所激起的框架中的框架元素体现出来。多个词元可以属于同一个框架,而且,具有多个义项的词语可以分别属于不同的框架。框架语义学将词语分领域、分框架的进行概括,并使用框架和框架之间的关系将框架建立成了一种语言学的本体网,为合理地解释词汇的语义信息提供了一种新的观点。

基于框架语义学的框架网络数据库中主要包含三个部分:框架库、词元库和句

子库。其中，框架库以框架为单位，对词语进行分类描述，并针对每个框架给出了明确的定义和框架下词元共有的框架元素，进而描述框架和其他框架之间的概念关系；词元库记录了词元的语义搭配模式和框架元素的句法实现方式；句子库记录带有框架语义标注信息的句子，即按照框架库所提供的框架和框架元素类型，标注句子的框架语义信息和句法信息，它可以作为训练数据供计算机处理语言使用。虽然，框架语义学及框架网络有着很高的研究价值和应用价值，但是，它的一些缺点经常受到很多研究者的质疑：

1. 框架的数量及覆盖率问题。框架库中究竟建立多少个框架就可以将语言中的所有词元完全覆盖，目前的框架对整个语言现象有多高的覆盖率。目前看来，这些问题没有一个明确的答案。但是，本文相信在受限领域中，框架个数是有限的，而且覆盖率也是比较高的。这一点在汉语框架网络中已经有了初步的例证，例如，汉语框架网络针对图书领域和旅游领域所建立的框架。
2. 框架网络中词元的关系不明确。FrameNet 中以框架为单位将词元进行聚类，但是并不关心单个词语之间的关系，例如“同义”和“反义”。当然，英文中，单个词语之间的关系可以参考 WordNet，而汉语中可以参考 HowNet，同义词词林等。

本文认为，任何的语义体系都不能完全覆盖所有的语言现象，不同的语义体系刻画了语言现象的不同侧面。不过，FrameNet 所具有的微小瑕疵并不能掩盖整个框架语义学体系和整个框架网络工程的耀眼光芒。

2.2 汉语框架网络介绍

汉语框架网络（CFN,Chinese FrameNet）工程是以 C.J.Fillmore 的框架语义学为理论基础、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。类似于 FrameNet，汉语框架网络知识库（CFN,Chinese FrameNet database）由框架库、句子库和词元库三部分组成。目前，CFN 已经对 2610 个词元构建了 230 个框架，标注了 15000 条句子，包含认知领域用词、科普文章常用谓词以及部分中国法律用词。

2.2.1 框架库

框架库对汉语的词汇按照所表示的活动场景（框架）的异同分类描述。框架库中包含框架名称、框架的定义、框架元素名称、框架元素的定义、框架之间的关系以及属于该框架的词元等。下面，本文以框架库中“陈述”框架来进行说明，框架定义

如表 2.1 所示。

表 2.1 框架库中“陈述”框架示例

框架名	陈述(Statement)
定义	说话者用语言向听话者传达信息的行为
核心框架元素	
媒介 Medium[Medium]	媒介是说话者用来传达信息的物理实体或渠道。
信息 Message[Msg]	说明说话者向听话者传达的内容，
说话者 Speaker[Spkr]	说话者是提供信息的人（无论是说还是写）。
非核心框架元素	
听话者 Addressee [Add]	信息传达的对象。
关涉 Topic[Top]	关涉是与信息有关的内容。
致因 Cause[Cau]	导致说话者说出信息的实体或事件。
程度 Degree[Degr]	描述事件有层级差别的属性。
形容 Depictive[Depic]	说话者在陈述中所处的状态。
修饰 Manner[Manr]	描述陈述活动的特征或所处状态的成分, 是对事件本身的描述。
时间 Time[Time]	说话者进行陈述的时间。
父框架	信息交流(Communication)
词元	评述(v) 报道(v) 说话(v) 描述(v) 综述(v) 演说(v) 声称(v) 等

从表 2.1 中可见,“陈述”框架主要是为了刻画说话者与听话者之间信息交流的行为。该框架中包含 10 个框架元素,其中,核心框架元素 3 个,非核心框架元素 7 个。“陈述”框架中的父框架为“信息交流”框架,并且“陈述”框架具有 87 个词元。

在汉语框架语义知识库(CFN)中,有一些框架元素(语义角色)可以出现在所有框架中,具有通用性,因此对这些框架元素进行单独描述和分析是非常有必要的,CFN 句子标注规范^[31]将它们定义为通用非核心元素,具体的标记集合如表 2.2 所示。

表 2.2 通用语义角色标记集

通用角色标记	含义	通用角色标记	含义
amount	物量	iter	动作次数
benef	受益人	loc_X	角色位置
circ	环境条件	malef	受损者
co_even	并行事件	manr	修饰
co_part	伴随者	mater	材料
contain_even	所属事件	mns	手段
cor_var	相关变量	part_iter	特殊重复

degr	程度	place	空间
depend	依据	rup	目的
depic	形容	rea	原因
instru	工具	reci	接受者
dur_action	动作时间量	role	身份
dur_fin_st	结果时间量	sco_role	角色范围
eve_desc	事件评价	subr	亚区
freq	频率	time	时间
time_role	角色事件时间		

2.2.2 词元库

词元库描述每一个词元的词义，并根据句子标注结果形成标注报告。在汉语中，很多词语都有多个义项。根据义项的不同，该词语作为一个词元被安放在不同的框架中。词元库中每个词元的记录主要由以下三部分内容组成：

1) 传统词典的词语释义：主要解释词元所激活的场景中有那些特有的参与者，这些参与者在活动中起什么作用等。

2) 句法实现方式报告：记录词元所支配的框架元素的句法特点。

3) 语义搭配模型汇总报告：记录词元与框架元素的组合方式。

上述三方面的内容本文不再具体解释，感兴趣的读者可以参考文献[32]。实际上，词元库的统计信息提供了大量的句法与语义的链接关系，这是当代的语言学家特别关注的问题。尤其是对于汉语来说，句法与语义之间的关系特别值得研究。

2.2.3 句子库

目前，CFN 的句子库中存储了 15000 多条带有框架语义标注信息的句子。CFN 句子库的句子标注语料来源于“121 篇阅读理解文本”和“北京大学 CCL 现代汉语语料库”。在为每个词元选取句子时，注意选择不同的语义搭配模型，并保证句子库的多样性。针对所选取的汉语句子，给定一个词元和该词元所属的框架，人工手动地给框架元素所在的成分标注上相应的框架元素名称、短语类型和句法功能三种信息。

下面是“陈述”框架下，词元“宣布”的一个标注例句：

例句 1: <spkr-np-subj 英 jn 方面 n> <eve_desc-vp-va 作为 v 报复 v 措施 n> 也 d <tgt=“陈述” 宣布 v> <msg-dj-obj 4 m 名 q 俄罗斯 nsy 大使馆 n 的 u 外交官 n 为 v 不 d 受 v 欢迎 v 的 u 人 n> 。 w <medium ini>

从上述例句可以看出，针对一个汉语句子，CFN 的标注需要包含如下几方面的内容：

(1)词元,指具有单一意义的词(或者词组、成语),也是语义角色标注过程中的目标词(tgt,target word)。例句1中的“宣布”即为本文所要标注的词元。

(2)框架,用来记录给定的词元所属于的框架。例句1中,词元“宣布”属于“陈述”框架。

(3)框架元素(FE,frame element),即句子中各个成分在目标词的所属框架中所充当的框架元素,即语义角色。如2.2.1节中给出的一样,框架元素分为核心框架元素、非核心框架元素和通用非核心框架元素三种。如果一些核心框架元素在句子中没有被实现,则可将其标注为零型式框架元素。例句1中,“spkr”和“msg”为“陈述”框架的核心框架元素,而“eve_desc”为通用非核心框架元素。由于句子中没有短语与语义角色“medium”对应,故将角色“medium”标注为零型框架元素。

(4)短语类型标记(PT,phrase type),即短语的句法属性类型,在CFN标注中“短语”是广义的,既指由两个或两个以上词语组成的结构,也包括由一个词语构成的句法单位。CFN的句子标注工作只标注框架元素所在的短语,目标词和句子中其他成分则不标注短语类型。

(5)句法功能标记(GF,grammatical function),CFN中的句法功能指的是所标短语相对于目标词的句法功能,即在短语和目标词所构成的句法关系中,该短语所起到的句法功能。

(6)零碎成分,主要包含目标词的支撑词(被标注为supp)和透明成分(被标注为null)两种。

本文仅研究对核心框架元素和非核心框架元素的标注,而不考虑通用非核心框架元素、短语类型等其它一些信息的标注。

2.3 CFN 句法语义分析模型的构建

CFN 句法语义分析的研究目标,是开发高性能的汉语句法语义分析器,包含现代汉语句子的句法结构分析器和词汇语义标注器两部分内容,为进行大规模真实文本的语义信息标注提供有力支持。

图2.1中给出了本文初步构建的CFN句法语义分析模型。输入一条汉语句子,CFN句法语义的分析主要包括三大阶段:

1. 词法、句法分析阶段:对未经任何标注的汉语句子进行词法分析和句法结构分析,为进一步分析汉语句子中的词元、词元的框架及相应的语义角色奠定基础。经过二十多年的探索,汉语的分词和词性标注技术趋近于成熟,但是汉语句法分析

性能，尤其是完全句法分析，仍然很不成熟。鉴于目前这方面的研究较多，本文不再给出具体模块的分解和实现。本文采用了正确的分词和词性标注结果。对于句法信息，本文中仅使用了自动分析得到的汉语的浅层句法分析信息，即：汉语基本块信息。

2. CFN 语义角色标注阶段：在词法、句法分析的基础上，确定汉语句子的目标词元及相应框架，并对该词元的语义角色信息进行详细分析。图 2.1 给出了识别 CFN 语义角色时的具体流程。下面针对每个模块进行具体分析：

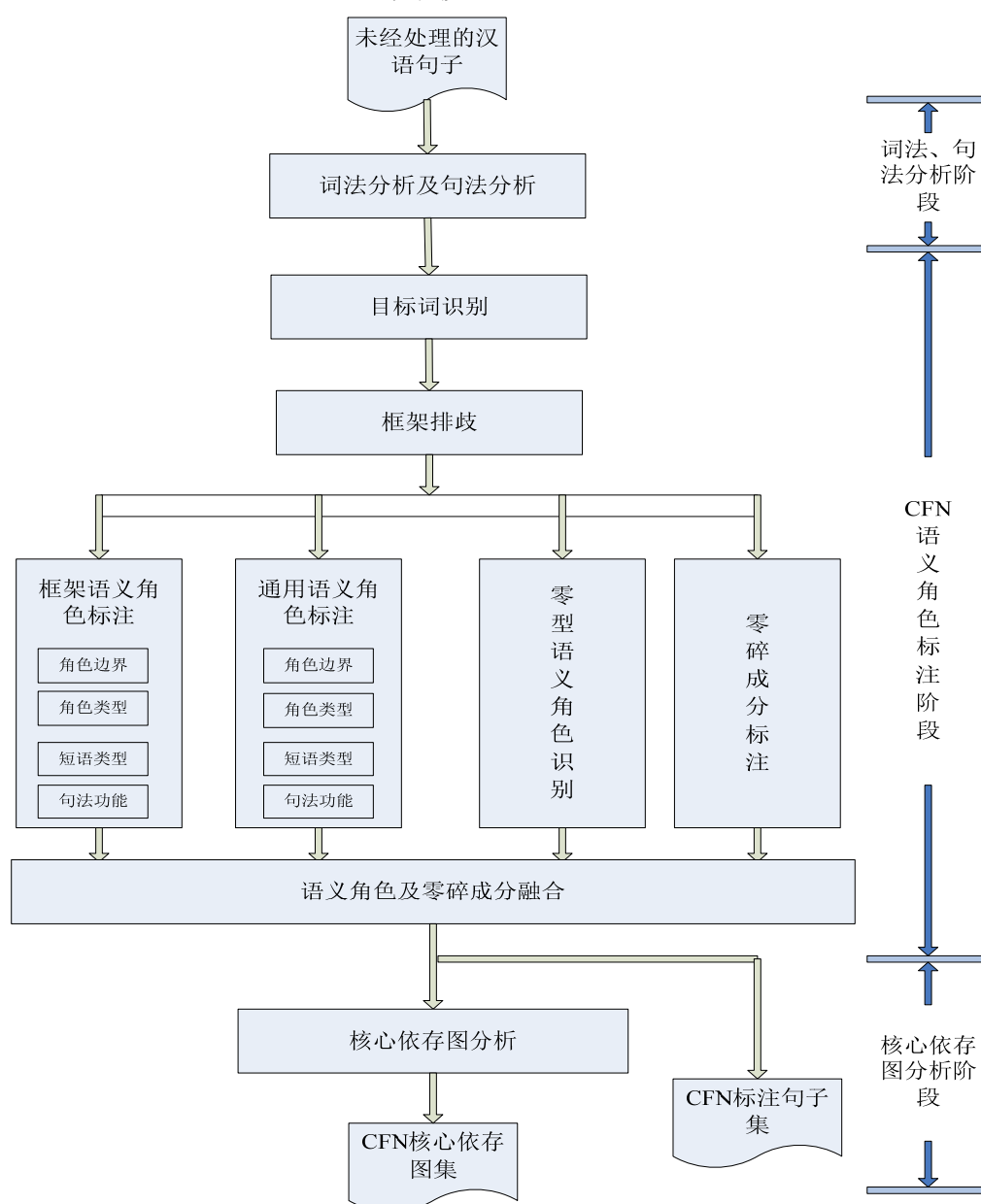


图 2.1 CFN 句法语义分析模型

(1) 目标词识别模块：识别出给定句子中可以激起框架的所有目标词元，主要

包含一些动词、名词、形容词和习用语。需要注意的是,单纯使用查找的方法在 CFN 词元库中进行搜索是一种拙略的实现办法。目前 CFN 中词元库中词元个数较少、覆盖率低,很多框架存在着词元缺失的情况。因此,“未登录词元识别”的研究是十分重要的。文献[33]中使用了支撑向量机将 WordNet 中的词语扩充到了英文 FrameNet 中,并人工选取了 100 个词元进行评价,发现准确率在 70%左右。是否可以将这种方法借鉴到 CFN 的未登陆词元识别中来是值得思考的。

(2) 框架排歧模块:给定目标词及其所在的句子,正确地识别该目标词所属的框架。表面上,这个任务类似与传统的词义排歧任务(WSD),但实际上却不相同。在 Semeval-2007 的评测任务中明显地体现出了这一点,这里不再过多解释。框架排歧任务可以被分解成两个级联的子任务。第一个是未知框架检测,主要目的是判断给定词元是否属于框架库中所给定的框架。这主要是针对目前框架库中框架缺失现象较严重而提出的。不过值得一提的是,在任何的词义体系中都存在着义项缺失的情况。目前对未知框架的检测主要使用一些异常点检测技术。文献[34]中便使用了 KNN 方法来判断一个词元的框架是否不同于框架库中给定的框架,从而推断出它是否激起了未知框架。关于异常点检测的技术还有很多,文献[35]和[36]中给出了这些方法的一些综述,这些方法均可以被借鉴到未知框架检测中来。当然,一类分类问题方面的研究也有很多,例如文献[37],这些方法也可以被在未知框架检测任务中使用。在这个任务的基础上,可以进一步进行“最近框架检测”任务,这里不再详述。第二个子任务是框架排歧,当多个框架中具有同一个词元并且句子中的该词元不属于未知框架时,必须正确地判断出该词元具体属于哪一个框架。这方面的文献较多,并且使用传统的分类方法可以解决。这里不再具体介绍。

(3) 语义角色标注模块:给定句子中的目标词及其所属框架,正确的识别出该词元的框架元素的边界和类型。具体包括图 2.1 中的框架语义角色标注模块、通用语义角色标注模块、零型语义角色识别模块、零碎成分标注模块和语义角色及零碎成分融合模块。框架语义角色标注模块主要针对框架中所定义的框架元素进行标注,主要分为角色边界识别、角色类型标注、角色的短语类型标注和句法功能标注;通用语义角色标注模块主要针对框架中未定义的通用语义角色进行标注,同样分为角色边界识别、角色类型标注、角色的短语类型标注和句法功能标注;零型语义角色识别模块主要识别句子中缺省的 NI 类型的框架元素;零碎成分识别模块主要识别句子中目标词的支撑成分和透明成分;语义角色及零碎成分融合模块是将上述四个模块的输出结果进行融合,从而形成句子的完整的 CFN 语义角色标注。本文认为零型

语义角色识别模块和零碎成分识别模块可以更多地使用一些启发式的规则进行处理。这四个模块之间不是相互独立的，可以有着相互的交互作用，图中没有给出相关的标识。在具体的标注实例中，存在着框架元素融合的现象，这可以被看作是机器学习中的“多标记学习”问题^[38]。在具体解决时，可以将语言学规则和相应的机器学习模型结合起来。需要进一步解释的是，图 2.1 中将框架语义角色标注和通用语义角色标注分开考虑主要从理论体系和学习技术两个方面来看的：

- 不同框架之间的通用语义角色是同质的。虽然通用语义角色刻画的是语义角色和目标词之间的关系，而讨论目标词必须给定框架。但是，不同框架下相同的通用语义角色表示的是同样的语义关系。例如，“manr”不管是在哪个框架下都表示谓词和论元之间的“修饰”关系。因此，将这些同质的语义角色放到一起进行标注有利于发现其中的规律，当然，前提是给定目标词元。
- 统计模型中将通用语义角色和框架语义角色分开标注有利于提升通用语义角色标注的性能。从文献[53]中可见，由于一个框架下通用语义角色的出现次数较少，导致条件随机场模型对通用语义角色的标注精度十分差，F 值仅达到 34%，而文献[81]中研究了通用语义角色标注任务，F 值达到 60%多。从实验结果来看，将通用语义角色单独标注有利于避免数据稀疏问题和标注模型的偏执问题，是值得尝试的一种方法。

本文仅针对框架语义角色标注模块的角色边界识别和角色类型识别进行了研究。本文中所给出的 CFN 框架语义角色标注任务的描述请参见 2.4 节，这里不再详述。

3. 核心依存图分析阶段：在给定同一个汉语句子的 CFN 语义角色标注的基础上，分析出汉语句子的核心依存图。英文 FrameNet 的核心依存图^[39]分析实质上是对一个给定句子中的谓词的部分分析，并需要确认以下内容：①在一套“控制项-依从项”的联系中，确认与控制项相关的框架，②确认框架的被表达或理解的框架元素，③确认表达框架元素的词汇核心，④确认框架元素句法表现的方式。英文 FrameNet 核心依存图的提出主要是为了得到句子中目标词的语义角色和句法实现之间的联系。它的提出为句子语义的部分形式化和语义推理打下坚实的基础。较英语来说，汉语具有更强的孤立语特点和意合特点，因此，汉语 CFN 的核心依存图描述体系不能完全照搬英文 FrameNet 的描述体系。举个简单的例子，句子“教地理的把那个学生给打了”，这个句子中词语“打”的施事“教地理的”如何进一步分析形成相应的核心依存图是值得探讨的。目前，汉语框架网络的研究还处在初步的阶段，

如何根据汉语的特点制定汉语的核心依存图描述体系是一件具有重大意义的研究工作。

针对汉语 CFN 的句法语义分析器构建的理论论证和实验论证目前已经踏出了坚实的一步。本文相信，随着对汉语 CFN 认识的深入，更好的句法语义分析模型可以被提出并实现。

2.4 汉语框架语义角色标注任务描述

考虑到汉语框架 CFN 的建设仍然属于初始阶段，可用的语料规模还比较小。为此，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，标出该目标词所支配的语义角色（框架元素，包括核心和非核心元素，不包含通用语义角色）。

本文从认知领域和其它领域选取 25 个框架，共 6723 条 CFN 标注句子进行了实验。考虑到目前 CFN 的全文标注数据较少，本文所选例句仅为词典编撰式的标注句子。随着语料规模的逐步扩大，本文将会深入开展核心依存图分析，话题的事件及其参与者链接及全文语义推理方面的研究。

2.5 本章小结

本章首先简要地介绍了英文 FrameNet 的发展情况。然后，对汉语框架网络目前的规模及框架库、句子库和词元库进行了简略介绍。在此基础上，提出了目前设想的 CFN 句法语义分析模型的概貌，其中，详细给出了 CFN 语义角色标注模块的具体流程。最后，本文概括出目前 CFN 语义角色标注任务的定义，并将它确定为本文的研究目标。

第3章 条件随机场模型介绍

本文将汉语框架语义角色标注看作是一种线性序列标注任务。条件随机场模型特别擅长于处理序列标注问题。因此，本文使用条件随机场模型来自动标注汉语框架语义角色。本章主要介绍了条件随机场模型，它可以很好地弥补隐马尔科夫模型、最大熵马尔科夫模型等一些有向图模型的缺陷。随后，本章分别针对条件随机场模型中参数估计、特征选取以及推理机制等一些细节方面的问题进行了详细地叙述。最后，本文简要介绍实验所使用的 CRF 工具包。

3.1 条件随机场模型

条件随机场模型（CRF，Conditional Random Fields）^[11]由 Lafferty 和 McCallum 等人于 2001 年提出。它将无向图中的团函数和最大熵框架有机地融合到一起，得到了一个用来解决序列标注和分割的概率模型。条件随机场模型不仅克服了隐马尔科夫模型^[40]的强独立性假设，而且不具有最大熵马尔科夫模型^[41]的标注偏执问题。因此，它可以显著地提高序列标注的精度。

条件随机场模型的原始定义如(3-1)所示。

$$P(Y|X) = \frac{1}{Z(X)} \exp\left(\sum_{e \in E, k} \lambda_k f_k(e, Y|_e, X) + \sum_{v \in V, k} \mu_k g_k(v, Y|_v, X)\right) \quad (3-1)$$

其中， X 为原始的输入序列， Y 为输出序列， $Z(X)$ 为归一化因子。 $f_k(\cdot)$ 和 $g_k(\cdot)$ 为相应的边特征函数和顶点特征函数， λ_k 和 μ_k 为特征函数 $f_k(\cdot)$ 和 $g_k(\cdot)$ 的参数。它们的具体含义请参考文献[11]，这里不再详细介绍。

作为条件随机场模型的一种特殊情况，线性条件随机场模型被广泛地使用到了汉语框架语义角色标注等线性序列标注任务。此时，输入序列 $X = x_1 x_2 \dots x_N$ ，输出序列 $Y = y_1 y_2 \dots y_N$ 。线性条件随机场模型的具体形式如(3-2)所示。

$$\begin{aligned} P(Y|X) &= \frac{1}{Z(X)} \exp\left(\sum_{t=1}^N \sum_k \lambda_k f_k(y_{t-1}, y_t) + \sum_{t=1}^N \sum_l \mu_l g_l(y_t, X)\right) \\ &= \frac{1}{Z(X)} \prod_{t=1}^N \exp\left(\sum_k \lambda_k f_k(y_{t-1}, y_t) + \sum_l \mu_l g_l(y_t, X)\right) \end{aligned} \quad (3-2)$$

其中， f_k 为输出序列 Y 中两个相邻时刻 y_{t-1} 和 y_t 的转移特征。 g_l 为输入序列 X 与输出序列 Y 中 t 时刻对应的输出节点 y_t 之间的顶点特征。 $Z(X)$ 为归一化因子，它

的函数形式如(3-3)所示

$$Z(X) = \sum_Y \prod_{t=1}^N \exp(\sum_k \lambda_k f_k(y_{t-1}, y_t) + \sum_l \mu_l g_l(y_t, X)) \quad (3-3)$$

线性条件随机场模型的求解会涉及如下三方面的问题：

问题 1：特征选择问题，即：选择哪些 f_k 和 g_l ？

问题 2：参数估计问题，即：给定所选的特征，怎样估计相应的权重参数 λ_k 和 μ_l ？

问题 3：模型概率推理问题，即：给定输入序列 X ，所选的特征函数集合以及相应的参数估计值，怎样高效地预测出相应的标注序列 Y^* ，使得 $Y^* = \arg \max_Y P(Y | X)$ ？

3.2 条件随机场模型的特征选取方法

不同于传统的分类问题，序列标注任务有着独特的特点。例如，序列标注模型中，一个序列中的每个标注单位有着较强的相关性，而序列和序列之间是独立的。很多研究者在进行序列标注时，假设序列中的每个元素之间是独立同分布的，并将序列标注任务看作是对每个序列中每个元素进行单点分类任务。在此基础上，使用最大熵或者支撑向量机模型进行训练。显然，这样的假设是不合理的。而条件随机场模型便考虑到了元素之间的相关性，并将一个序列看作是一个样本。关于这些特点的详细分析，请参见文献[42]。正是由于序列之间各个元素具有相关性，开窗口技术^[43]才被用来对序列标注进行特征选择。

目前，序列标注的特征选择方法主要有如下四种：

1. 对候选特征空间中每个特征进行单个约减。文献[55]中，使用似然增益对候选特征集合中的单个特征进行选择。实验结果表明，该方法可以有效地降低特征空间的维数，从而改善模型的预测精度。但是，这种方法针对单个特征进行，时间复杂度较高，并且对特征的解释性不好。例如，这种方法为什么仅保留了一部分词特征，而舍弃了另一部分词特征，很难作出解释。另外，每次都贪婪地选择似然增益最大的特征，这样做难以得到全局最优的特征子集。
2. 主观经验选取法。组块分析任务中，很多研究者直接将窗口固定进行实验，例如，文献[21]将每类特征的窗口设定为[-2,2]。这种做法固然简单，但是具有很大的盲目性，选择到最优特征窗口的概率较小。

3. 贪婪选取法。为了避免主观经验选取的盲目性，很多学者对特征类型进行排序，并逐个贪婪地搜索使得模型性能达到最优的特征窗口。文献[44]和文献[45]中便使用了这种方法。这种方法类似于传统实验设计中的单因素轮换法，很难找到最优的特征模板。而且，进行第 n 个实验需要知道第 $n-1$ 个实验的结果，只能以串行的方式搜索最优窗口，实验的时间比较长。
4. 正交表选择法。使用正交表进行特征选择的基本思想可以参考文献[46]，它将每种类型的特征看作是正交表中的因素，将每类特征的候选窗口看作是相应的水平，那么把各种特征的各种窗口的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[47]中有详细的描述。但是，这类方法有两个较明显的缺陷：第一，当一类特征的窗口为多个（大于2个）时，要求各个窗口水平之间要有包含关系。第二，当特征类型比较多时，要求相应的正交表也比较大。

本文中采用正交表的方法对汉语框架语义角色标注模型所使用的特征模板进行选择。实验结果表明，基于正交表的特征选择方法可以避免主观选取的盲目性和贪婪算法的局部最优性，能够有效地改善模型的预测性能。

3.3 条件随机场模型的参数估计方法

给定特征函数的集合，估计这些特征函数所对应的权重参数 $\Theta = \langle \lambda_1, \dots, \lambda_M, \mu_1, \dots, \mu_L \rangle$ 是条件随机场模型训练时必须解决的问题。本节给出 CRF 模型中权重参数的两种估计方法：极大似然估计和最大化后验估计。其中，最大化后验估计是为了解决极大似然估计中的过学习问题而产生的。

3.3.1 条件随机场模型权重参数的极大似然估计

给定一组观测样本 $D = \{ \langle X, Y \rangle^{(1)}, \langle X, Y \rangle^{(2)}, \dots, \langle X, Y \rangle^{(N)} \}$ ，条件随机场模型的对数似然函数可以写成(3-4)的形式。

$$\begin{aligned}
 L_{MLE}(\Theta) &= \log \left(\prod_{\langle X, Y \rangle \in D} P(Y | X)^{\tilde{p}(X, Y)} \right) \\
 &= \frac{1}{N} \sum_{i=1}^N \left(\sum_t \left(\sum_k \lambda_k f_k(y^{(i)}_{t-1}, y^{(i)}_t) + \sum_l \mu_l g_l(y_t, X^{(i)}) \right) - \log Z(X^{(i)}) \right)
 \end{aligned} \tag{3-4}$$

其中， $y^{(i)}_t$ 为输出序列 $Y^{(i)}$ 在时刻 t 的标记。

权重参数 Θ 的估计应该是使(3-4)达到最大的极大似然估计。然而，快速地找到权重参数的极大似然值并不容易，得到条件随机场模型权重参数的极大似然估计需要使用迭代的方法进行求解。其中，比例迭代算法(iterative scaling)^[48]是一种传统的训练算法，被使用到最大熵模型的训练过程中。当然，它也适用于 CRF 模型的训练过程。不过，目前的研究结果表明，一些拟牛顿算法，例如 L-BFGS 算法^[49]，在求解参数时表现得更为高效。

3.3.2 条件随机场模型的极大似然估计缺陷

条件随机场模型的极大似然估计的存在性和唯一性方面的证明至今仍未见到相关的文献。作为条件随机场的理论基础，最大熵模型的极大似然估计的存在性和唯一性方面的证明可以参考文献[50]和文献[51]。文献[50]中定义了数据点可分条件，并且证明了当最大熵训练数据中每两类数据点都是完全可分时，参数的极大似然估计不存在。文献[51]将这个结论推广到了其它的模型中。

文献[52]详细叙述了最大熵模型的极大似然估计的过拟合现象。它的阐述是从 n-gram 语言模型入手的。为了缓解过拟合现象，该文献将高斯平滑技术引入到了最大熵的参数估计方法中。可以证明，加入高斯平滑后，最大熵模型的参数估计存在并且唯一。

大量的实验表明，条件随机场模型的极大似然估计也容易产生过拟合现象。最直接的一个证据是，使用条件随机场模型进行封闭测试，实验结果的 F 值可以达到近 100%。但是，开放测试的 F 值却严重退化。

另外，很多实验结果^[53]表明，在样本不均衡的情况下，条件随机场模型的预测精度并不理想。这一点在最大熵模型中也有所体现，但是，这个现象并没有给出理论方面的证明。文献[54]中仅给出了样本不均衡对两项 LOGISTIC 模型预测性能的影响。如何将这个理论证明及结论推广到最大熵模型和条件随机场模型上，还有待进一步思考。

3.3.3 条件随机场模型参数的最大化后验估计

条件随机场模型极大似然参数估计的过拟合现象有很多方法可以改进。目前最常用的一类方法是强迫权重参数服从一定的参数分布或约束，例如，高斯分布^[55]和拉普拉斯分布^[56]。加入参数的分布条件后的参数估计被称作极大后验估计(MAP, maximum a posteriori)。

在条件随机场的极大似然函数中加入文献[55]所介绍的高斯分布可以使得权重

参数的估计值服从均值为 0，方差为 σ^2 的高斯分布。参数估计表现出一定的稠密性，即，靠近 0，但不精确为 0。而加入文献[56]所介绍的拉普拉斯分布后，参数估计表现出了稀疏性，即，大量的参数为 0。本文尝试使用加入拉普拉斯分布的条件随机场模型进行汉语框架语义角色标注实验，实验结果并不理想，这主要是由于大量的特征参数为 0，从而影响了模型的性能。

为了实现和使用的方便，很多工具包假设所有参数服从的分布是相同的，但实际上，不同类型的特征对应的参数分布可能也不同，更不可能仅仅是高斯分布和拉普拉斯分布。目前，已有相应的方法可以强迫模型参数介于高斯先验和拉普拉斯先验之间。这方面的研究有很多文献，具体可以参考文献[57]。但是，这些模型是否适用于自然语言处理的相关任务需要进一步的实验和理论方面的论证。

3.4 条件随机场模型的概率推理机制

CRF 模型的预测问题主要关心给定输入序列 X ，所选的特征函数集合以及相应的参数估计值，怎样高效地预测出相应的标注序列 Y^* ，使得 $Y^* = \arg \max_Y P(Y | X)$ ？其中，归一化因子 $Z(X)$ 的计算和最优序列 $Y^* = \arg \max_Y P(Y | X)$ 的寻找是需要重点解决的两个问题。

3.4.1 归一化因子的计算公式

假设 CRF 模型所使用的转移特征向量和顶点特征向量为 $\langle f_1, \dots, f_M \rangle$ 和 $\langle g_1, \dots, g_L \rangle$ ，它们所对应的权重参数向量分别为 $\langle \lambda_1, \dots, \lambda_M \rangle$ 和 $\langle \mu_1, \dots, \mu_L \rangle$ ，那么，令

$$B(y_{t-1}, y_t) = \begin{cases} \exp(\sum_{k=1}^M \lambda_k f_k(y_{t-1}, y_t)) & 1 < t \leq N \\ 1 & t = 1 \end{cases} \quad (3-5)$$

$$U(y_t, X) = \exp(\sum_{l=1}^L \mu_l g_l(y_t, X)) \quad (3-6)$$

则(3-2)可以被改写为(3-7)所示的公式。

$$P(Y | X) = \frac{1}{Z(X)} \prod_{t=1}^N B(y_{t-1}, y_t) U(y_t, X) \quad (3-7)$$

在此基础上，本文分别定义前向概率^[58] $\alpha_t(y_t)$ 和后向概率^[58] $\beta_t(y_t)$ 如下：

$$\alpha_t(y_t) = \begin{cases} U(y_t, X) & t = 1 \\ \sum_{y_{t-1}} \alpha_{t-1}(y_{t-1}) B(y_{t-1}, y_t) U(y_t, X) & 1 < t \leq N \end{cases} \quad (3-8)$$

$$\beta_t(y_t) = \begin{cases} U(y_t, X) & t = N \\ \sum_{y_{t+1}} \beta_{t+1}(y_{t+1}) B(y_t, y_{t+1}) U(y_t, X) & 1 < t \leq N \end{cases} \quad (3-9)$$

则有(3-10)成立:

$$Z(X) = \sum_{y_1} \beta_1(y_1) = \sum_{y_N} \alpha_N(y_N) = \sum_{y_t} \frac{\alpha_t(y_t) \beta_t(y_t)}{U(y_t, X)} \quad (3-10)$$

因此, 对于归一化因子 $Z(X)$ 的计算, 本文只需要根据定义递归地计算出序列 X 中某一时刻 t 对应的所有可能的 y_t 的前、后向概率的值, 并使用(3-10)进行相应的求解即可。

首先引入边际概率来刻画输入序列 X 在时刻 t 输出 y_t 的概率。

定义: 给定输入序列 $X = x_1 x_2 \dots x_N$, 则 CRF 模型在时刻 t 输出标记 y_t 的边际概率为:

$$P(y_t | X) = \sum_{y_1 y_2 \dots y_{t-1} y_{t+1} \dots y_N} P(Y | X) \quad (3-11)$$

下面给出 CRF 模型中的归一化因子计算公式(3-10)的证明。

定理: 给定输入序列 $X = x_1 x_2 \dots x_N$, 归一化因子 $Z(X)$ 可以使用(3-12)进行计算。

$$Z(X) = \sum_{y_1} \beta_1(y_1) = \sum_{y_N} \alpha_N(y_N) = \sum_{y_t} \frac{\alpha_t(y_t) \beta_t(y_t)}{U(y_t, X)} \quad (3-12)$$

其中, $\alpha_t(y_t)$ 和 $\beta_t(y_t)$ 分别为 CRF 模型在 t 时刻输出 y_t 的前向概率和后向概率。

证明: 由前向概率 $\alpha_t(y_t)$ 和后向概率 $\beta_t(y_t)$ 的定义可知:

$$\begin{aligned}
 \alpha_t(y_t) &= \sum_{y_{t-1}} \alpha_{t-1}(y_{t-1}) B(y_{t-1}, y_t) U(y_t, X) \\
 &= \sum_{y_1} \sum_{y_2} \cdots \sum_{y_{t-1}} U(y_1, X) \cdots U(y_t, X) \cdot B(y_1, y_2) \cdots B(y_{t-1}, y_t) \\
 &= U(y_t, X) \cdot \sum_{y_1 y_2 \cdots y_{t-1}} B(y_{t-1}, y_t) \prod_{i=1}^{t-1} U(y_i, X) B(y_{i-1}, y_i)
 \end{aligned} \tag{3-13}$$

同理，对于后向概率 $\beta_t(y_t)$ ，有：

$$\beta_t(y_t) = U(y_t, X) \cdot \sum_{y_{t+1} \cdots y_N} \prod_{i=t+1}^N U(y_i, X) B(y_{i-1}, y_i) \tag{3-14}$$

故，将(3-13)和(3-14)带入到(3-12)的右边可以得到：

$$\begin{aligned}
 \sum_{y_t} \frac{\alpha_t(y_t) \beta_t(y_t)}{U(y_t, X)} &= \sum_{y_t} \frac{\alpha_t(y_t) (U(y_t, X) \cdot \sum_{y_{t+1} \cdots y_N} \prod_{i=t+1}^N U(y_i, X) B(y_{i-1}, y_i))}{U(y_t, X)} \\
 &= \sum_{y_t} [U(y_t, X) \cdot \sum_{y_1 y_2 \cdots y_{t-1}} B(y_{t-1}, y_t) \prod_{i=1}^{t-1} U(y_i, X) B(y_{i-1}, y_i)] \cdot [\sum_{y_{t+1} \cdots y_N} \prod_{i=t+1}^N U(y_i, X) B(y_{i-1}, y_i)] \\
 &= \sum_{y_1 y_2 \cdots y_N} \prod_{t=1}^N B(y_{t-1}, y_t) U(y_t, X) \\
 &= Z(X)
 \end{aligned}$$

故等式 $Z(X) = \sum_{y_t} \frac{\alpha_t(y_t) \beta_t(y_t)}{U(y_t, X)}$ 成立。

当 $t=1$ 或 N 时，便有 $Z(X) = \sum_{y_1} \beta_1(y_1) = \sum_{y_N} \alpha_N(y_N)$ 成立。

定理证毕。

实际上，将(3-13)和(3-14)带入到边际概率的定义(3-11)中，可以得到

$$P(y_t | X) = \frac{\alpha_t(y_t) \beta_t(y_t)}{Z(X) U(y_t, X)}。 \text{ 由于 } \sum_{y_t} P(y_t | X) = 1, \text{ 故 } Z(X) = \sum_{y_t} \frac{\alpha_t(y_t) \beta_t(y_t)}{U(y_t, X)}。$$

3.4.2 最优标注序列的寻找

对于最优输出序列 Y^* ，Viterbi 算法^[59]可以提高使用动态规划技术来提高寻找的效率。在 Viterbi 算法中，最基本的思想是使用一个概率 δ 来标记最优可能的路径。在 CRF 模型中，Viterbi 算法的详细过程如下：

1)初始化：对于所有的 y_1 ，设 $\delta_1(y_1) = U(y_1, X)$ ，且 $\Delta_1(y_1) = 0$ ；

2)递归计算输入序列 X 中每个标注单位 x_t 所对应的输出标记 y_t 的最佳路径:

$$\delta_t(y_t) = [\max_{y_{t-1}} \delta_{t-1}(y_{t-1}) B(y_{t-1}, y_t)] U(y_t, X)$$

$$\Delta_t(y_t) = \arg \max_{y_{t-1}} [\delta_{t-1}(y_{t-1}) B(y_{t-1}, y_t)] U(y_t, X)$$

3)到达输入序列 X 的最后一个标注单位 x_N 时, 计算这个标注单位的最佳输出标记:

$$P^* = \max_{y_N} [\delta_N(y_N)] \quad y_N^* = \arg \max_{y_N} (\delta_N(y_N))$$

4)从 x_N 的最佳输出标记开始, 依次向前取得每个标注单位的最佳输出标记:

$$y_t^* = \Delta_{t+1}(y_{t+1}^*) \quad t = N-1, N-2, \dots, 2, 1$$

使用上述的 Viterbi 算法找到的标注序列 $Y = y_1 y_2 \dots y_N$ 便是输入序列 $X = x_1 x_2 \dots x_N$ 的最优标注序列。

在对 CRF 的应用研究中, 本文发现命题“给定输入序列 $X = x_1 x_2 \dots x_N$, 存在标注序列 $Y = y_1 y_2 \dots y_N$, 使得 $Y^* = \arg \max_Y P(Y | X)$, 当且仅当对于任意的 $t \in [1, N]$, 使得 $y_t^* = \arg \max_{y_t} P(y_t | X)$ ”在很多情况下成立。该命题的详细证明有待于进一步研究。

3.5 IOB2 标注策略及不合理序列的处理

实际上, 目前对线性序列处理的任务主要分为两种。第一种为序列标注, 经典的例子是对分好词的句子进行词性标注, 其中每个词语都具有一个相应的词性标记。另外一种为序列切分 (Sequence segmentation), 即, 识别出序列中有用的元素子序列, 并对该子序列赋予一个相应的标记。传统的组块分析大多属于这种任务, 本文的汉语框架语义角色标注也不例外。目前处理序列切分的可行方法是使用一定的描述策略, 将序列切分任务转化为序列标注任务。常用的描述策略主要有两大类^[60]: 第一大类为 Inside/Outside 描述策略, 第二大类为 Start/End 描述策略。这里不再详细介绍。

本文使用 IOB2 策略来刻画每个句子的框架语义角色标注序列。IOB2 策略属于 Inside/Outside 描述策略中的一种。当然, 使用其它的描述策略进行实验也是十分有必要的。在 IOB2 策略中, 标记 B 代表语义角色块的开始, I 代表语义角色块的延续,

O 代表块外词。本文中，将汉语框架语义角色的边界和类别进行识别，因此所使用的 IOB2 策略的标记集合为 $\text{TagSet} = \bigcup_{x \in \text{FESet}} \{B-x, I-x, O\}$ ，其中 FESet 是一个框架下的所有框架元素的标记集合。具体的例子在第四章给出，这里不再赘述。

虽然训练的时候，IOB2 策略可以将所有的样本正确地进行表述。但是，测试的时候，有可能产生一些不合理的 IOB2 序列，导致无法组合形成相应的汉语框架语义角色块。本文提出如下两种方法对不合理序列进行过滤和校正。

第一种方法是直接修改 CRF 模型的标注序列。针对“... O I-X ...”或“...B-X I-Y ...”或“...B-X I-X I-Y...”等一些不合理的输出序列，将直接将检测到的第一个不合理的标记转化成新的块，具体的算法如下：

设 $S=(s[1],s[2],...,s[n])$ 为当前句子的标注序列；

For i=1 to n

 If $s[i]='I-X'$ and ($s[i-1]='O'$ or $s[i-1]='B-Y'$ or $s[i-1]='I-Y'$) then

$s[i]='B-X'$

 Endif

Endfor

另外，本文将目标词所对应的输出标记强制设为 O。

第二种方法是将条件随机场的自动输出设置为 $Y^* = \arg \max_{Y \wedge C(Y)=true} P(Y | X)$ ，其中，

$C(Y)$ 表示序列 Y 应该满足的约束集合。本文中，约束集合主要有如下两个：

(1) 输出序列合法：对于两层的标注集合，本文不允许标注序列中出现“... O I-X ...”或“...B-X I-Y ...”或“...B-X I-X I-Y...”等无法正确形成语义角色块的标记存在。

(2) 目标词所对应的输出标记必须为“O”。

本文采用第一种方法对条件随机场模型的自动预测结果进行修正。相关实验证明，在非法序列较少的情况下，两种方法的校正对模型性能的提升没有显著的差异。

3.6 条件随机场模型的改进

自 2001 年条件随机场被提出至今，很多研究者提出了条件随机场模型的改进方法，有些是针对条件随机场模型的时间复杂度进行的，而更多的是将条件随机场进行扩展和变型以适应相应的任务。

针对条件随机场模型时间复杂度方面的改进主要有两种方法。文献[61]使用错误

驱动的方法对条件随机场模型中的参数进行估计，参数更新时仅使用了加法运算。因此，这种模型可以有效地改善条件随机场模型训练时间过长的缺陷。但是，相应的实验结果表明，这种方法的性能较传统的条件随机场模型有所退化。文献[62]将 Beam Search 的方法引入到条件随机场模型的推理机制中，并使用 KL 距离来近似原来的推理过程，这种方法有效地降低了条件随机场模型的训练时间。

自然语言处理中的很多任务可以被看作是多个序列标注的级联模型，例如，本文的汉语框架语义角色自动标注任务便是典型的一种。在这种级联模型中，首先需要进行分词和词性标注，然后，进行短语的组块分析，最后进行汉语框架语义角色的识别。其中，每种序列的标注又需要使用前一种标注序列的信息。使用纯粹的级联方式会使得错误在模型中不断积累。为此，文献[63]将这些输出看作是一个矩阵，矩阵的每一行代表不同层面的序列信息，将原有条件随机场模型扩展到输入为单个序列，输出为多个序列的情况中，并将其命名为动态条件随机场 (Dynamic Conditional Fields Model)。

由于信息抽取任务中，很多序列之间存在长依存性和一些全局特征，因此，很多文献使用不同的方法将全局特征和长依存特征引入到条件随机场模型中。文献[64]利用了 Gibbs Sampling 的方法，采用模拟退火(simulated annealing)算法将“标记一致性”和“抽取模板一致性”两类全局特征融合入 CRF 模型中。实验结果表明，融合入全局特征的实验有较明显的提高。文献[65]也给出了一种两阶段的模型将全局特征和长依存特征融合到条件随机场模型中。在汉语框架语义角色标注中，有哪些全局特征或者长依存特征存在，以及如何将其融合到条件随机场模型中，可以从上述文献中找到相应的思路。

针对层次型文本数据的标注问题，文献[66]提出了一种树状的条件随机场模型，以便于进行语义标注。而文献[67]提出了一种层次条件随机场，并将其使用到了网页文本的信息抽取任务中。汉语框架的核心依存图任务抽取是否可以使用这种方法实现是一件值得考虑的事情。

3.7 条件随机场工具包介绍

随着对条件随机场研究的深入，很多关于条件随机场模型的工具包被实现。比较出名的有，CRF++^[68]，CRF Suite^[69]，FlexCRF^[70]和 MALLET^[71]。其中，CRF++和 CRF Suite 两个工具包都实现了 L1 惩罚条件随机场模型和 L2 惩罚条件随机场模型。CRF++工具包中还实现了 CRF 的感知器训练方法，而 CRF Suite 使用纯 C 语言实现，

训练速度比较快。FlexCRF 实现了二阶转移的条件随机场模型，并且还提供了一个并行的条件随机场工具包。MALLET 实现了传统的条件随机场模型和动态条件随机场模型。

本文使用了 CRF++ 工具包进行实验，该工具包实现了 CRFL1、CRFL2 以及 Perceptron 三种方法。其中，CRFL1 和 CRFL2 方法都是原来的经典算法，它们的唯一区别在于平滑项的选取上。而 Perceptron 算法在 CRF 的训练过程中，使用了神经计算中一种经典的错误驱动算法。在进行参数估计时，该工具包使用了 L-BFGS 算法来迭代寻找 MLE 的最大值。本文使用了 CRFL2 算法，并选取 $C=1$ 进行参数平滑。考虑到输出序列的一阶转移特征在大量的实验中表现出了优越的性能，本文将这种特征作为必选特征，实验部分不再具体声明。

CRF++ 工具包的使用方法参考文献[68]，这里不再具体介绍。但是，在使用时，需要注意如下几方面：

(1) 该工具包使用模板文件将数据文件扩展成多个特征。每个特征的具体特征值仅有 0 和 1 两种。该工具包仅能使用示性函数作为特征，而不支持其它的实值分段函数。因此，怎样将条件随机场模型的特征函数扩展为实值函数，序列标注中使用示性函数是否已经足够等问题需要进一步实验和理论论证。

(2) 该工具包中使用的 L-BFGS 算法中，设置的 Hessian 矩阵使用单位阵近似。这样做主要是出于计算效率上的考虑。是否可以使用其它对角阵或一般矩阵来替换单位阵，哪些矩阵会显著降低参数估计的均方误差等问题需要进一步思考。

(3) 使用该工具包自带的命令行进行测试仅能得到概率最大的输出序列 $Y^* = \arg \max_Y P(Y | X)$ ，而无法得到加入约束的最优序列 $Y^* = \arg \max_{Y \wedge C(Y)=true} P(Y | X)$ 。如果需要得到后面的最优序列，需要使用该工具包 SDK 目录下中提供的 crfpp.h 文件和 libcrfpp.lib 文件，具体如何调用可以参考 example.cpp 文件。本文不再给出具体例子。

3.8 本章小结

本章首先介绍了条件随机场的定义和理论框架，并给出了线性条件随机场的具体形式。随后的几节分别针对线性条件随机场的参数估计、特征选择和推理机制三方面的问题进行了论述，并总结了其中存在的一些值得思考的问题。尤其是在特征选择方面，本文使用了正交表的方法。该方法在传统的试验设计领域已经具有完善的理论，但是，在自然语言处理领域却比较罕见。序列标注的正交表特征选择方法的理论论证是一件十分有意义的课题。

本章随后简要地介绍了 IOB2 策略，它可以有效地将序列分割问题转化为序列标注问题。但是，测试阶段的自动预测序列中有可能出现不合理的 IOB2 序列。针对这个问题，本文提出了两种处理方法。

为了进一步了解条件随机场模型的发展趋势，本文介绍了其它学者针对条件随机场改进所做的一些工作。本章的最后给出了条件随机场模型的一些工具包，并着重地针对本文所使用的工具包进行了介绍。

第4章 汉语框架语义角色标注的基线模型

本章使用词语作为语义角色的标注单位，提取词层面的特征，并使用条件随机场模型建立了汉语框架语义角色自动标注系统。为后续基本块信息的加入，提供了一个基本的参照结果。

4.1 基于词的汉语框架语义角色标注模型

4.1.1 模型的形式化描述

汉语框架语义角色标注可以被看作是一个标注单位为词语的序列标注问题。针对一个汉语句子，给定标注的词元和它所属的框架，针对该框架所具有的框架元素集合 $FESet$ ，本文通过设立一个标记集合 $TagSet = \bigcup_{x \in FESet} \{B-x, I-x, O\}$ ，并对句子中的每个词语标注一个合适的标记 $y \in TagSet$ ，从而实现汉语框架语义角色的自动标注。标注集合中，除 O 外，每个元素有两部分组成。第一部分代表词语在语义角色块中的位置，语义角色块起始位置的词语用 B 标记，内部位置的词语用 I 标记。第二部分代表词语所在语义角色块的类型。对于语义角色块外的词语，统一使用 O 进行标记。例句(1)的语义角色块的 IOB2 序列标记如下（为清晰起见，省略了词性标记）。

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

通过这样的方法，汉语框架语义角色标注就被转化为了一个序列标注问题。因此，本文可以使用条件随机场模型对汉语框架语义角色标注进行处理。在下一节中，本文详细给出条件随机场模型所使用到的特征集合及特征的选取方法。

4.1.2 模型标注范围的选取

对于汉语框架语义角色的标注，建立模型时只能针对给定的一个框架进行。这主要是因为不同的框架的框架元素是不同的，只有在同一个框架下的词元才具有相同的框架元素。虽然目前 CFN 句子库中已经人工标注了 15000 条句子，但是，每个框架的平均句子数不足 200 条。因此，汉语框架语义角色的自动标注研究仍然面临着严重的数据稀疏问题。

Gildea 和 Jurafsky 在 2002 年进行的 FrameNet 语义角色标注的开创性研究工作^[27]中，将每个框架下的每个词元看做是不同的模型。给定框架，针对框架中的每个词

元，他们使用该词元的所有标注例句作为训练集，建立相应的标注模型。他们认为，不同的目标词元具有不同的语义角色搭配模式。但是，这种做法使得数据稀疏问题变得更加明显。

而张国华的研究工作^[72]使用一个框架下所有的例句作为训练集，使用条件随机场对汉语框架语义角色标注进行建模。这种做法虽然充分利用了一个框架下的所有的例句，但是，他并没有考虑到不同的词元的语义角色搭配模式的差异。

针对一个框架下所有的词元，本文认为词性不同的词元具有的语义角色的分布及相应搭配模式是不同的。本文从“陈述”框架下的所有句子中分别抽取出了动词词元的语义角色搭配模式和名词词元的语义角色搭配模式。由于篇幅有限，这里不再具体给出。从语义角色的分布情况来看，名词词元很少与语义角色“msg”搭配，而和“top”搭配较多；动词词元恰好相反。而对于语义角色搭配模式来说，名词词元绝大多数语义角色搭配模式中目标词出现在模式最后面，而这种情况在动词的语义角色搭配模式中却相对较少。

表 4.1 中分别给出了“陈述”框架下动词词元的例句和名词词元的例句中各个语义角色出现的频次。

表 4.1 “陈述”框架下动词词元和名词词元的语义角色分布对比

语义角色	动词词元(1381 条例句)		名词词元(253 条例句)	
	角色个数	所占比例	角色个数	所占比例
spkr	833	23.66%	157	30.60%
msg	1302	36.98%	9	1.75%
medium	540	15.34%	47	9.16%
degr	2	0.06%	6	1.17%
time	273	7.75%	32	6.24%
manr	205	5.82%	117	22.81%
depic	57	1.62%	1	0.19%
cau	3	0.09%	0	0.00%
top	42	1.19%	144	28.07%
add	264	7.50%	0	0.00%
总计	3521	100%	513	100%

图 4.1 给出了动词词元和名词词元的各语义角色分布的对比情况。

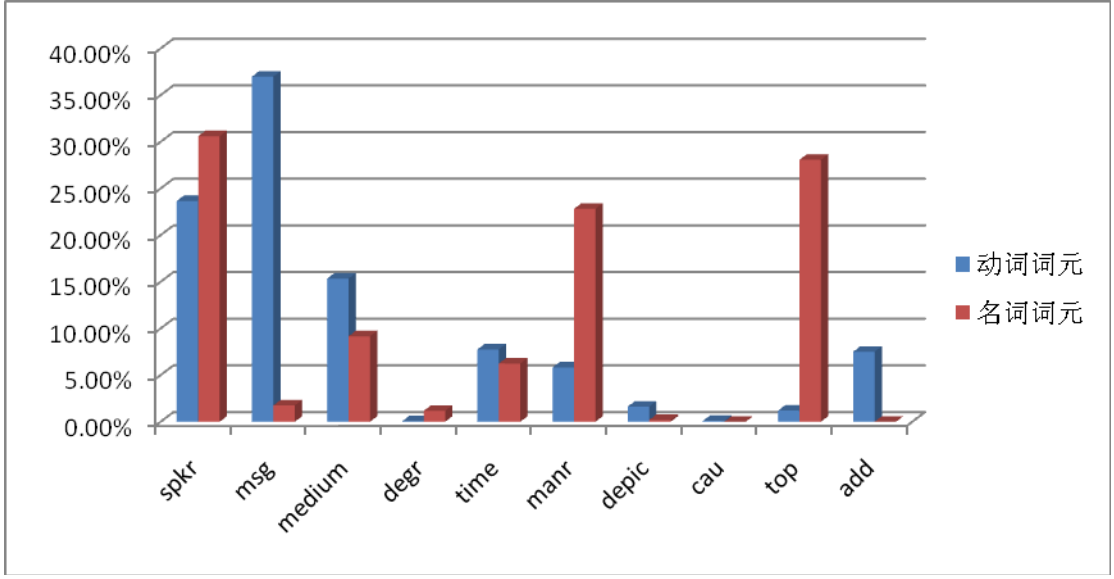


图 4.1 动词词元和名词词元中各语义角色分布的对比

从表 4.1 和图 4.1 看出, "陈述"框架的动词与名词的语义角色分布差异比较显著。判断不同词性的词元之间的分布的显著差异, 本文可以借助如下定义的 χ^2 检验。

假设某框架有 K 种语义角色, 有 M 种词元词性, 设第 m 中词性的第 k 种语义角色的频次为 C_{km} , 其中, $m=1\dots M, k=1\dots K$, 则语义角色的分布矩阵 D 如(4-1)所示。

$$D = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1M} \\ C_{21} & C_{22} & & C_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & & C_{KM} \end{pmatrix} \quad (4-1)$$

则可以通过如(4-2)所示的 χ^2 统计量来判断第 m 种词性的语义角色分布是否显著不同于所有词性的平均语义角色分布。

$$\chi^2_m(K-1) = \sum_{k=1}^K \frac{\left(C_{km} - \frac{\left(\sum_{k=1}^K C_{km} \right) \cdot \left(\sum_{m=1}^M C_{km} \right)}{\sum_{k=1}^K \sum_{m=1}^M C_{km}} \right)^2}{\frac{\left(\sum_{k=1}^K C_{km} \right) \cdot \left(\sum_{m=1}^M C_{km} \right)}{\sum_{k=1}^K \sum_{m=1}^M C_{km}}} \quad (4-2)$$

上面给出的 χ^2 检验实际上是看第 m 类词性词元的语义角色分布是否显著不同于平均的分布。这里, 本文设显著水平为 95%, 也就是, 本文以 0.95 的概率来保证

第 m 类词元词性的语义角色分布与平均的分布差异的显著性。针对表 4.1 中给出的“陈述”框架下的动词和名词的语义角色分布，可以计算得到，“陈述”框架动词的检验值 $\chi^2_v = 143.6357 > \chi^2(9) = 16.919$ ，“陈述”框架名词的检验值 $\chi^2_n = 985.5705 > \chi^2(9) = 16.919$ 。因此，“陈述”框架的动词词元和名词词元的语义角色分布情况是显著不同的。

通过 χ^2 检验，本文可以将一个框架下的 M 种词元词性分成显著和不显著两种。对于显著的词元，如果句子库中相应的句子较少，仍然不适于单独训练形成模型。根据显著性和词元词性所具有的句子数目，本文按照如下的方法形成模型：

(1) 将显著并且句子数大于语义角色个数 20 倍的每种词元词性单独训练形成一个模型。

(2) 将句子数小于语义角色个数 20 倍的和显著的所有词元词性共同训练成一个模型。

其中“20 倍”的确定仅仅是一种经验的取值。

使用上述的模型划分方法，本文将“陈述”框架划分成动词词元模型和名词词元模型两种。

4.2 特征描述及选取方法

本章所研究的基于词的基线系统中，本文假设语料库中的汉语句已经经过了正确的分词、词性标注，并且已经识别出正确的目标词和相应的框架信息。从给定的这些信息中，本文可以总结出如下几种特征：

词特征：指当前所要标注的词条。

词性特征：指当前所要标注的词条的词性。

位置特征：指当前标注的词条相对于目标词的位置。

目标词特征：指当前标注的词条所在句子的目标词。

词特征使用当前所要标注的词语作为特征。但是，词特征具有稀疏性，也就是说，训练集中不可能出现所有的词条，进行开放测试时，很可能出现训练集中没有的词特征。尤其是当训练集比较小时，词特征的稀疏性表现得更为严重。此时，模型的泛化能力也就变差。

对于词性特征，本文使用正确的词性标记作为特征。例句库中的词性标注遵循的是山西大学的 973 词性规范（具体见附录 A）。

位置特征刻画了所要标注的每个词与句子中给定目标词之间的位置关系。对于位置特征，本文使用标记集合 $\{L, R, T\}$ 进行刻画。其中，标记“L”代表当前标注的

词语在目标词的左边;标记“R”代表当前标注的词语在目标词的右边;标记“T”代表当前标注的词语是目标词。位置特征可以为寻找特定的语义角色提供简单的信息。一般来说,对于“陈述”框架下的动词词元,语义角色“spkr”一般出现在目标词的左边,而语义角色“msg”一般出现在目标词的右边。对于“陈述”框架下的名词词元,语义角色一般都出现在目标词的左边。

目标词特征指汉语句子里中激起框架的具体的词条。一个汉语句子里的不同的词所具有的目标词特征是相同的。一个汉语框架所具有的语义角色少则几个,多则几十个,但不是所有的语义角色都会同时出现在汉语句子里。给定该框架下的一个例句,一般来说,在该句子里同时出现的语义角色不会超过六个(PropBank 体系的结论),并且语义角色出现及相应的搭配模式与该句中的目标词的关系是十分重大的。例如,当汉语句子的目标词为“告诉 v”并且激起“陈述”框架时,该汉语句子里一般都会出现“spkr”、“add”和“msg”,语义角色的搭配模式是“spkr(说话者)+告诉+add(听话者)+msg(信息)”。目标词特征的引入可以很大程度上体现该目标词的语义角色的搭配模式。不过,值得注意的是,目标词特征也具有稀疏性。如果测试集中遇到了训练集中没有的目标词特征,那么,模型就很难根据目标词特征来判断句子的语义角色及搭配方式。

在语义角色标注模型中,决定当前词语的语义角色标记除了当前词本身所具有的信息外,很大程度上还依赖于它相邻的词的信息。在模型中,本文可以使用开窗口的方式来对当前词所处的环境进行扩展。借助于开窗口的方式,本文还可以针对每种特征引入相应的二元组特征和三元组特征,以及不同特征之间的搭配特征。针对上述提到的四种特征,本文可以考虑如下几种组合特征和搭配特征:

- (1) 相邻词的二元组合特征;
- (2) 相邻词词性的二元组合特征;
- (3) 相邻词的位置的二元组合特征;
- (4) 相邻词的位置的三元组合特征;
- (5) 词与其词性的搭配特征;
- (6) 词与位置的搭配特征;
- (6) 词性与词的位置的搭配特征;
- (7) 词与目标词的搭配特征;

特征窗口大小的确定也是模型中特征选择的重要环节。窗口不能定义的太大,否则将加剧特征的稀疏程度。在基线模型中,本文将窗口定在 3 以内,那么可以供

选择的主要特征列表如表 4.2 所示。

表 4.2 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）	[0, 0]			

表 4.2 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。

从表 4.2 中可以看出，针对词，本文考虑了四种窗口大小的选取方式：① $[0, 0]$ 代表只选取当前词作为特征；② $[-1, 1]$ 代表选取当前词，及其前一个词，后一个词作为特征；③ $[-2, 2]$ 代表选取当前词，及其前一个词，后一个词，前第二个词，后第二个词做为特征；④ $[-3, 3]$ 代表选取当前词，及其前一个词，后一个词，前第二个词，后第二个词，前第三个词，后第三个词作为特征。当前词的词性和词的位置特征的窗口大小的选择也是这四种。这三个特征为基本特征，所以窗口最小为 $[0, 0]$ ，也就是说这三个特征为必选。目标词为必选特征，也就是说，在每个训练模板中都要有。

针对词的二元组合，本文也分别考虑了四种选取方式：①不选取(-)；② $[-1, 1]$ 表示选取当前词左右各一个标注单位的词所形成的二元组合特征；③ $[-2, 2]$ 表示选取左右各两个标注单位的词所形成的二元组合特征；④ $[-3, 3]$ 表示选取左右各三个标注单位的词所形成的二元组合特征；这里需要注意的是，当窗口为 $[0, 0]$ 时，由于窗口中只容纳一个标注单位，无法形成二元组合特征，等同于不选取的情况，因此，词的二元组合特征窗口没有 $[0, 0]$ 的情形。对词性、词的位置情况类似。

本文仅考虑了位置特征的三元组合特征，这是因为位置特征的标记集合元素数目小（仅有三个），三元组合特征的搭配情况也少，不会使特征稀疏。

对于词/词性、词/位置和词性/位置这三种两两搭配特征,本文分别考虑了四种选取方式:①不选取(-);②选取当前标注单位[0, 0]的搭配特征;③选取左右各一个标注单位[-1, 1]的搭配特征;④选取左右各两个标注单位[-2, 2]的搭配特征;

对于词和目标词的搭配特征,本文仅考虑两种选取方式:①不选取(-);②选取当前标注单位[0, 0]的搭配特征;

上述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,显然不现实。本文退而求其次,在一批适量的候选的特征模板中,挑出最好的模板。这一批适量的候选的模板应满足:1)数量不能太多,因为计算量的原因;2)要有"代表性",也即在 $4^{10} \times 2$ 种全部特征模板中具有一定"代表性"的模板。传统的做法是逐个往模型中加入特征,根据经验设置每个特征的窗口大小,从中选出一个较好的模板,这种对特征窗口大小"盲目性"的设置,很可能会使得CRF模型失去其优点。本章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[47]中有详细的描述。此处对应的就是11个因子(特征),前10个因子每个因子4个水平(窗口大小),最后一个因子为2水平(窗口大小),文献[47]中可以选用的正交表有 $L_{32}(4^9 \times 2^4)$,为了用这个正交表,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。使用这个正交表,只需按表里列出的各组合(每种组合对应一个模板),选择32个具有良好代表性的特征模板,做训练、测试,从中选择一个较好的模板就可,大大降低了计算量。32个模板的选取标准,完全依据测试集上语义角色标注的F值的大小,F值大者为优。从实验结果可以看出,正交表方法选出的模板比一般的用经验"试"出的特征模板要好。

4.3 语料选取

由于在建立语义角色标注模型时主要采用了经验主义的做法,因此,所选训练和测试语料中句子的数目直接影响到实验结果。为了能够得到比较可信的实验结果,本文从130个框架中挑选出25个框架作为汉语框架语义角色标注实验语料。这25个框架来自于"认知"领域和其它领域。表4.3给出了这25个框架名称,它们所含有的框架元素的个数以及拥有例句的数目。

表 4.3 从语料库中选出的框架分布情况表

框架名	框架元素个数	核心元素个数	非核心元素个数	例句个数	词元个数
感受	6	4	2	569	61

知觉特征	5	1	4	345	36
思想	3	3	0	141	15
联想	5	4	1	185	19
自主感知	14	4	10	499	27
查看	9	2	7	320	31
思考	8	2	6	283	33
非自主感知	13	3	10	379	28
获知	9	4	5	258	15
相信	8	2	6	218	26
记忆	12	2	10	298	17
包含	6	2	4	126	24
宗教信仰	5	2	3	54	4
因果	7	4	3	140	14
陈述	10	3	7	1634	87
拥有	4	2	2	170	15
适宜性	4	3	1	70	7
发明	12	2	10	198	13
计划	6	2	4	90	9
代表	7	3	4	80	8
范畴化	11	4	7	125	10
证明	9	2	7	101	11
鲜明性	9	2	7	260	30
外观	10	4	6	106	13
属于某类	8	3	5	74	7
合计	200	69	131	6723	560

4.4 语料切分

针对一个框架下的所有例句，首先按照词元词性不同和每种词性例句的数目的多少，将其划分成不同的模型。具体的划分方法见 4.1.2 小节。针对所划分的每种模型下的所有例句，本文按照 5: 5 的比例将其切分成训练集和测试集。切分时，本文尽量使每种语义角色在训练集和测试集中都出现，并且比例尽量接近 1:1。

表 4.4 给出了“陈述”框架中训练集和测试集的语义角色分布对比，可以看出训练集合测试集中例句的数目以及各个元素的数目的比例基本上接近 1: 1。

表 4.4 “陈述”框架中训练集和测试集的语义角色分布对比

框架 元素	动词词元(1381 句)				名词词元(253 句)			
	训练集(691 句)		测试集(690 句)		训练集(126 句)		测试集(127 句)	
spkr	415	23.54%	417	23.56%	81	30.80%	76	30.40%
msg	655	37.15%	647	37.20%	5	1.90%	4	1.60%
medium	260	14.75%	280	14.67%	23	8.75%	24	9.60%

degr	1	0.06%	1	0.06%	0	0.00%	6	2.40%
time	140	7.94%	133	7.98%	19	7.22%	13	5.20%
manr	104	5.90%	101	5.95%	60	22.81%	57	22.80%
depic	32	1.82%	25	1.87%	1	0.38%	0	0.00%
cau	2	0.11%	1	0.11%	0	0.00%	0	0.00%
top	28	1.59%	14	1.47%	74	28.14%	70	28.00%
add	126	7.15%	138	7.13%	0	0.00%	0	0.00%
合计	1763	100%	1757	100%	263	100%	250	100%

其它 25 个框架中训练集和测试集的例句数和语义角色的个数不再具体给出。

4.5 评价指标

正确识别一个语义角色块指的是语义角色块的边界正确，并且语义角色块的类型也识别正确。为此，本文使用如下的准确率(Precision)，召回率(Recall)和 F 值(F-Score)来评价汉语框架语义角色标注模型的性能。

假设模型标注出的语义角色块数为 C_p ，其中正确的功能块（左右边界正确，且语义角色类型正确）数目为 C_c ，测试集中的语义角色块的数目为 C_o ，那么，准确率

$$P = \frac{C_c}{C_p}; \text{ 召回率 } R = \frac{C_c}{C_o}; \text{ F 值 } F = \frac{2PR}{P+R}。$$

为了从多方面来评价汉语框架语义角色标注模型的性能，本文分别给出如下几方面的指标：

- (1) 每种类型的语义角色块标注的准确率、召回率和 F 值。
- (2) 核心语义角色整体标注的准确率、召回率和 F 值。
- (3) 非核心语义角色整体标注的准确率、召回率和 F 值。
- (4) 所有语义角色标注的准确率、召回率和 F 值。

4.6 特征重要性分析

客观地分析模型中每种特征的重要性是十分重要的。在语义角色标注的相关文献中，每类特征重要性的分析采用的方法^[73]一般是在基线模型 A 的基础上，加入某类特征 X 进行实验得到模型 B，然后对模型 A 和 B 进行卡方检验，来得到特征 X 的重要程度（是否显著）。但是，这种方法最大的缺陷在于，单独加入特征 X 的显著性不能刻画它在最优模型中的显著程度。

不同于上述的分析方法，本文使用正交表的极差分析方法对模型中每种特征的重要程度进行分析。极差分析方法的理论具体请参照文献[47]，这里不再赘述。针对模型所使用的正交表的每列的特征，本文分别给出了它的最大水平、最大均值、最

小水平、最小均值和极差（最大均值与最小均值之差），并按照极差的大小对各个特征进行了排序。极差越大的特征的重要程度越大，反之亦然。当然，这种分析方法也具有一些缺陷。对于未进入正交表的特征（如，目标词特征），极差分析无法给出它的重要性度量。获得了每个特征的极差和最大均值所对应的水平后，可以使用正交表的“算一算”的方法进一步来提升模型的性能。由于这种提升不是必然的，因此，本文不再给出“算一算”的实验结果。

另外，当模型的特征没有填满所使用的正交表时，采用正交表方差分析的方法来得到每种特征。本文没有采用这种方法对各个特征进行分析。

4.7 实验结果

根据表 4.2 中特征模板的正交表选择方法，需要进行 32 组实验。对于测试集，在例句库中已标注好的分词、词性标注的基础上，进行语义标注的测试，得到每个模板的测试结果，从 32 个模板中挑选一个 F-值最大的模板作为模型的最优模板。

本节先给出“陈述”框架的实验的详细结果及相应的分析，最后给出 25 个框架的每个的测试结果和整体测试结果。需要说明的是：以下在每个框架的测试集上的结果，均以测试例句的已有分词和已有的词性标注为基础，得到的语义角色的标注结果。

4.7.1 “陈述”框架动词词元的实验结果

表 4.5 给出了“陈述”框架动词词元的实验结果。从表 4.5 中可以看出，模板 1 的实验的 F1 值最好，F1-值达到 69.22%。而基于传统的经验“试”出来的模板选取方法，得到的最好的 F1-值为 66.3%^[74]，与正交表模板选取方法有显著的差别。

表 4.5 “陈述”框架动词词元模型的实验结果

模板 编号	非法 序列	特征数	训练时间	核心框架语义角色			非核心框架语义角色			所有框架语义角色		
				P	R	F	P	R	F	P	R	F
1	3	785780	536.20s	75.47%	69.12%	72.16%	77.02%	46.25%	57.79%	75.73%	63.75%	69.22%
21	5	1660660	565.25s	76.38%	67.86%	71.87%	80.00%	45.52%	58.02%	76.98%	62.61%	69.05%
31	5	1262660	636.38s	76.02%	68.15%	71.87%	76.76%	44.79%	56.57%	76.14%	62.66%	68.75%
12	3	794880	699.81s	75.14%	68.15%	71.48%	76.89%	46.73%	58.13%	75.44%	63.12%	68.73%
0	1	2547500	539.00s	76.17%	67.78%	71.73%	80.18%	43.10%	56.06%	76.80%	61.98%	68.60%
6	5	1228860	574.73s	75.50%	68.08%	71.60%	77.41%	44.79%	56.75%	75.81%	62.61%	68.58%
9	3	1390380	759.36s	76.63%	67.34%	71.68%	78.85%	43.34%	55.94%	76.99%	61.70%	68.50%
17	1	2621580	644.97s	76.11%	67.34%	71.46%	81.08%	43.58%	56.69%	76.90%	61.75%	68.50%
20	2	1663420	686.92s	75.80%	66.89%	71.07%	81.17%	43.83%	56.92%	76.65%	61.47%	68.22%
26	4	1703560	600.63s	75.58%	67.49%	71.31%	77.83%	43.34%	55.68%	75.94%	61.81%	68.15%

10	2	2214420	591.58s	75.93%	67.11%	71.25%	79.11%	43.10%	55.80%	76.43%	61.47%	68.14%
23	4	1884500	528.44s	75.97%	67.04%	71.23%	78.17%	43.34%	55.76%	76.33%	61.47%	68.10%
27	2	3448260	449.86s	75.98%	66.59%	70.98%	81.00%	43.34%	56.47%	76.77%	61.13%	68.06%
18	3	1290760	674.36s	76.17%	66.37%	70.93%	80.72%	43.58%	56.60%	76.90%	61.01%	68.04%
11	2	800620	642.45s	74.90%	67.26%	70.87%	76.23%	45.04%	56.62%	75.12%	62.04%	67.96%
29	3	1730560	527.31s	74.81%	67.19%	70.80%	79.13%	44.07%	56.61%	75.50%	61.75%	67.94%
4	3	1394560	630.72s	75.40%	67.26%	71.10%	78.12%	42.37%	54.95%	75.83%	61.41%	67.86%
22	3	2550200	625.78s	76.08%	66.96%	71.23%	79.33%	39.95%	53.14%	76.56%	60.61%	67.66%
25	3	2159420	589.95s	75.89%	66.29%	70.77%	78.03%	42.13%	54.72%	76.23%	60.61%	67.53%
15	4	2967060	670.09s	75.91%	64.96%	70.01%	80.93%	42.13%	55.41%	76.70%	59.59%	67.07%
13	4	1821400	583.00s	75.49%	66.44%	70.68%	76.28%	39.71%	52.23%	75.61%	60.16%	67.00%
28	3	2348500	564.14s	76.00%	66.22%	70.78%	76.08%	38.50%	51.13%	76.01%	59.70%	66.88%
7	2	2495340	508.39s	75.04%	65.77%	70.10%	79.81%	40.19%	53.46%	75.76%	59.76%	66.82%
30	2	1741620	405.19s	75.02%	65.25%	69.80%	78.70%	41.16%	54.05%	75.60%	59.59%	66.65%
2	1	2674780	614.00s	74.70%	65.03%	69.53%	79.15%	40.44%	53.53%	75.38%	59.25%	66.35%
16	2	720200	772.05s	73.46%	65.70%	69.36%	72.76%	43.34%	54.32%	73.34%	60.44%	66.27%
8	6	2670880	408.41s	74.68%	64.73%	69.35%	79.90%	40.44%	53.70%	75.47%	59.02%	66.24%
19	1	2419960	579.80s	75.00%	63.84%	68.97%	82.52%	41.16%	54.93%	76.15%	58.51%	66.17%
3	5	1741840	393.42s	74.70%	64.36%	69.14%	77.73%	41.40%	54.03%	75.18%	58.96%	66.09%
14	5	2178600	473.20s	74.52%	63.32%	68.46%	78.67%	42.86%	55.49%	75.20%	58.51%	65.81%
24	5	443180	584.59s	72.70%	64.21%	68.19%	72.88%	41.65%	53.00%	72.73%	58.91%	65.09%
5	2	1938940	427.72s	74.33%	61.83%	67.51%	82.97%	36.56%	50.76%	75.54%	55.89%	64.25%

表 4.6 中给出了模板 1 所对应的各种特征的选取方式。模板 1 所对应的详细实验结果见表 4.7。从表 4.7 可以看出,动词词元模型中核心语义角色的识别精度较高,F 值达到 72.16%,而非核心语义角色的识别精度却仅达到 57.79%,明显低于总体水平。这表明当训练样本不均衡时,条件随机场模型的训练偏执严重影响了非核心语义角色的识别性能。

表 4.6 “陈述”框架动词词元语义角色标注的最优模板

特征类型	选取方式	解释
词	$[-1, 1]$	当前词以及其左右各一个词
词的二元组合	-	不选取
词性	$[0, 0]$	当前词的词性
词性的二元组合	$[-2, 2]$	当前词与左右各两个词的相邻词词性的二元组合
位置	$[0, 0]$	当前词的位置
位置的二元组合	$[-2, 2]$	当前词与左右各两个词的相邻词位置的二元组合
词/词性	$[-1, 1]$	当前词以及左右各一个词的词和词性的二元搭配
词/位置	$[0, 0]$	当前词的词和位置的二元搭配
词性/位置	$[-2, 2]$	当前词与左右各两个词的词性与位置的二元搭配
目标词	$[0, 0]$	当前词所在句子的目标词
位置三元组	-	不选取

词/目标词

-

不选取

表 4.7 动词模型的详细实验结果

名字	原来	预测	正确	P	R	F
medium	280	221	190	85.97%	67.86%	75.85%
msg	647	647	458	70.79%	70.79%	70.79%
spkr	417	363	281	77.41%	67.39%	72.05%
核心元素合计	1344	1231	929	75.47%	69.12%	72.16%
depic	25	17	5	29.41%	20.00%	23.81%
degr	1	0	0	0.00%	0.00%	0.00%
manr	101	39	30	76.92%	29.70%	42.86%
add	138	99	92	92.93%	66.67%	77.64%
top	14	6	1	16.67%	7.14%	10.00%
time	133	87	63	72.41%	47.37%	57.27%
cau	1	0	0	0.00%	0.00%	0.00%
非核心元素合计	413	248	191	77.02%	46.25%	57.79%
合计	1757	1479	1120	75.73%	63.75%	69.22%

4.7.2 “陈述”框架名词词元的实验结果

表 4.8 给出了“陈述”框架名词词元的 32 组实验结果。从中可以看出模板 21 的实验结果最好，总体 F1-值达到 57.58%。

表 4.8 “陈述”框架名词词元模型的实验结果

模板 编号	非法 序列	特征数	训练时间	核心框架语义角色			非核心框架语义角色			所有框架语义角色		
				P	R	F	P	R	F	P	R	F
21	2	307692	32.00s	79.07%	32.69%	46.26%	77.67%	54.79%	64.26%	78.08%	45.60%	57.58%
26	2	313530	32.81s	76.19%	30.77%	43.84%	75.24%	54.11%	62.95%	75.51%	44.40%	55.92%
6	1	253358	28.66s	75.56%	32.69%	45.64%	71.82%	54.11%	61.72%	72.90%	45.20%	55.80%
12	3	155106	35.50s	68.00%	32.69%	44.16%	68.91%	56.16%	61.89%	68.64%	46.40%	55.37%
29	2	313348	29.19s	71.74%	31.73%	44.00%	69.91%	54.11%	61.00%	70.44%	44.80%	54.77%
18	1	243726	31.55s	73.33%	31.73%	44.30%	73.08%	52.05%	60.80%	73.15%	43.60%	54.64%
23	1	360808	28.81s	76.19%	30.77%	43.84%	70.27%	53.42%	60.70%	71.90%	44.00%	54.59%
0	1	430654	27.47s	78.05%	30.77%	44.14%	72.38%	52.05%	60.56%	73.97%	43.20%	54.55%
9	3	227290	35.73s	68.89%	29.81%	41.61%	73.58%	53.42%	61.90%	72.19%	43.60%	54.36%
11	3	167314	31.06s	73.91%	32.69%	45.33%	66.67%	53.42%	59.32%	68.71%	44.80%	54.24%
31	1	233464	30.55s	69.57%	30.77%	42.67%	70.91%	53.42%	60.94%	70.51%	44.00%	54.19%
1	2	163380	25.25s	66.67%	30.77%	42.11%	71.56%	53.42%	61.18%	70.06%	44.00%	54.05%
4	2	234864	35.16s	69.77%	28.85%	40.82%	71.96%	52.74%	60.87%	71.33%	42.80%	53.50%
10	3	378098	29.25s	74.42%	30.77%	43.54%	70.48%	50.68%	58.96%	71.62%	42.40%	53.27%
22	2	412902	35.41s	75.00%	28.85%	41.67%	70.75%	51.37%	59.52%	71.92%	42.00%	53.03%

25	1	387030	32.98s	75.00%	28.85%	41.67%	71.84%	50.68%	59.44%	72.73%	41.60%	52.93%
20	2	281386	33.13s	73.17%	28.85%	41.38%	71.84%	50.68%	59.44%	72.22%	41.60%	52.79%
7	1	420714	27.44s	76.32%	27.88%	40.85%	74.23%	49.32%	59.26%	74.81%	40.40%	52.47%
3	1	310660	22.17s	76.19%	30.77%	43.84%	72.63%	47.26%	57.26%	73.72%	40.40%	52.20%
17	1	427364	31.78s	76.32%	27.88%	40.85%	72.73%	49.32%	58.78%	73.72%	40.40%	52.20%
2	0	452998	35.02s	79.49%	29.81%	43.36%	71.88%	47.26%	57.02%	74.07%	40.00%	51.95%
13	1	306348	28.72s	69.05%	27.88%	39.73%	70.48%	50.68%	58.96%	70.07%	41.20%	51.89%
19	0	422898	29.38s	69.23%	25.96%	37.76%	77.42%	49.32%	60.25%	75.00%	39.60%	51.83%
14	1	382508	25.67s	74.36%	27.88%	40.56%	74.47%	47.95%	58.33%	74.44%	39.60%	51.70%
28	1	387324	30.38s	69.77%	28.85%	40.82%	68.87%	50.00%	57.94%	69.13%	41.20%	51.63%
8	1	463554	23.20s	74.36%	27.88%	40.56%	73.68%	47.95%	58.09%	73.88%	39.60%	51.56%
27	3	593208	28.16s	75.68%	26.92%	39.72%	73.20%	48.63%	58.44%	73.88%	39.60%	51.56%
16	1	128744	39.08s	63.27%	29.81%	40.52%	66.97%	50.00%	57.25%	65.82%	41.60%	50.98%
30	1	306096	23.28s	71.43%	28.85%	41.10%	70.41%	47.26%	56.56%	70.71%	39.60%	50.77%
15	0	521094	36.39s	71.79%	26.92%	39.16%	69.61%	48.63%	57.26%	70.21%	39.60%	50.64%
24	3	66752	31.34s	50.85%	28.85%	36.81%	67.96%	47.95%	56.22%	61.73%	40.00%	48.54%
5	1	316764	24.20s	77.78%	26.92%	40.00%	70.00%	43.15%	53.39%	72.22%	36.40%	48.40%

表 4.9 给出了模板 21 所对应的特征组合方式, 详细的实验结果在表 4.10 中给出。从表 4.10 中可以看出, 非核心语义角色的标注精度要明显高于核心语义角色的标注精度, 这主要是因为非核心语义角色的训练比较充分导致的。

表 4.9 “陈述”框架名词词元语义角色标注的最优模板

特征类型	水平	解释
词	[-1,1]	当前词与左右各一个词
词的二元组合	-	不选取
词性	[-1,1]	当前词与左右各一个词的词性
词性的二元组合	[-2,2]	当前词与左右各两个词的相邻词的词性的二元组合
位置	[-1,1]	当前词与左右各一个词的二元组合
位置的二元组合	[-1,1]	当前词与左右各一个词的相邻词位置的二元组合
词/词性	[-2,2]	当前词与左右各两个词的词和词性的二元搭配
词/位置	[-2,2]	当前词与左右各两个词的词和位置的二元搭配
词性/位置	[0,0]	当前词的词性和位置的二元搭配
目标词	[0,0]	当前词所在句子的目标词
位置三元组	[-2,2]	当前词与左右各两个词的两个词位置的三元组合
词/目标词	[0,0]	当前词与目标词的二元搭配

表 4.10 名词模型的详细实验结果

名字	原来	预测	正确	P	R	F
medium	24	4	2	50.00%	8.33%	14.29%
msg	4	0	0	0.00%	0.00%	0.00%
spkr	76	39	32	82.05%	42.11%	55.65%

核心元素合计	104	43	34	79.07%	32.69%	46.26%
depic	0	0	0	0.00%	0.00%	0.00%
degr	6	0	0	0.00%	0.00%	0.00%
manr	57	52	40	76.92%	70.18%	73.39%
add	0	0	0	0.00%	0.00%	0.00%
top	70	47	37	78.72%	52.86%	63.25%
time	13	4	3	75.00%	23.08%	35.29%
cau	0	0	0	0.00%	0.00%	0.00%
非核心元素合计	146	103	80	77.67%	54.79%	64.26%
合计	250	146	114	78.08%	45.60%	57.58%

4.7.3 “陈述”框架的语义角色标注实验结果

表 4.11 “陈述”框架语义角色标注的最终实验结果

框架名称	核心框架语义角色			非核心框架语义角色			所有框架语义角色		
	P	R	F	P	R	F	P	R	F
陈述 v	75.47%	69.12%	72.16%	77.02%	46.25%	57.79%	75.73%	63.75%	69.22%
陈述 n	79.07%	32.69%	46.26%	77.67%	54.79%	64.26%	78.08%	45.60%	57.58%
陈述 (1)	75.59%	66.51%	70.76%	77.21%	48.48%	59.56%	75.94%	61.48%	67.95%
陈述 (2)	74.89%	65.89%	70.10%	75.76%	50.83%	60.84%	75.10%	61.51%	67.63%

表 4.11 中,“陈述(1)”这一行是按“陈述 v”和“陈述 n”的最好实验结果的语义块统计得到的结果;“陈述(2)”这一行表示不区分动词、名词例句,放在一起训练、测试得到的结果。从表 4.11 可以明显看出,对名词词元的标注的结果比较差,特别是核心元素的标注召回率只有 32.69%,总 F1-值只有 46.26%,与动词的 F1-值相差 25.90%。虽然将动词、名词放到一起训练、测试的结果的 F1-值为 67.63%,与单独对动词训练、测试的 F1-值为 69.22%只有 1.59%的差距,但是测试集中名词词元的例句较少,相应的语义块也少(250/1757),对整体的标注结果的影响没有明显体现出来。因此,按词元词性分开训练、测试对名词词元的标注是有益,是有必要的。

4.8 实验分析

对于语义角色标注模型的实验结果,本文从极差分析、性能分析和性能标注曲线分别给出相应的结论。

4.8.1 特征的极差分析

本文针对动词模型和名词模型分别给出了相应的极差分析,具体的细节见表 4.12 和表 4.13。从表 4.12 中可以看出,特征“词性/位置”的极差最大,为 1.59%,是动词模型中最重要的特征。而表 4.13 所示的名词模型中,特征“词的二元组”的

极差达到了 2.39%，是名词模型中最重要的特征。

表 4.12 “陈述”框架动词模型的极差分析

编号	特征类型	均值最大			均值最小			极差
		水平	窗口	均值	水平	窗口	均值	
1	词性/位置	4	[-2, 2]	68.06%	1	-	66.48%	1.59%
2	词的二元组	1	-	68.18%	4	[-3, 3]	66.88%	1.30%
3	位置二元组	4	[-3, 3]	68.16%	1	-	66.91%	1.25%
4	词/词性	4	[-2, 2]	67.74%	1	-	66.74%	1.00%
5	词/位置	4	[-2, 2]	67.78%	1	-	66.92%	0.86%
6	位置	4	[-3, 3]	67.87%	1	[0, 0]	67.05%	0.82%
7	词性二元组	3	[-2, 2]	67.79%	1	-	67.16%	0.63%
8	词	2	[-1, 1]	67.75%	1	[0, 0]	67.24%	0.51%
9	位置三元组[-1, 1]	1	-	67.62%	2	[0, 0]	67.27%	0.35%
10	位置三元组[-2, 0]	2	[0, 0]	67.57%	1	-	67.33%	0.24%
11	词/目标词	2	[0, 0]	67.55%	1	-	67.34%	0.21%
12	词性	2	[-1, 1]	67.57%	4	[-3, 3]	67.35%	0.21%
13	位置三元组[0, 2]	2	[0, 0]	67.53%	1	-	67.36%	0.17%

表 4.13 “陈述”框架名词模型的极差分析

编号	特征类型	均值最大			均值最小			极差
		水平	窗口	均值	水平	窗口	均值	
1	词的二元组	1	-	54.30%	4	[-3, 3]	51.37%	2.93%
2	词/词性	4	[-2, 2]	54.02%	1	-	51.73%	2.29%
3	词性/位置	4	[-2, 2]	53.77%	1	-	51.52%	2.24%
4	词/位置	4	[-2, 2]	53.78%	1	-	52.19%	1.59%
5	位置二元组	4	[-3, 3]	53.48%	1	-	52.07%	1.41%
6	位置	4	[-3, 3]	53.31%	1	[0, 0]	51.98%	1.33%
7	词	3	[-2, 2]	53.39%	1	[0, 0]	52.37%	1.02%
8	词性	2	[-1, 1]	53.52%	4	[-3, 3]	52.62%	0.90%
9	位置三元组[-1, 1]	1	-	53.14%	2	选取	52.73%	0.41%
10	位置三元组[0, 2]	2	选取	53.12%	1	-	52.74%	0.38%
11	词性二元组	4	[-3, 3]	53.16%	1	-	52.79%	0.36%
12	位置三元组[-2, 0]	2	选取	53.00%	1	-	52.87%	0.12%
13	词/目标词	2	选取	52.96%	1	-	52.91%	0.04%

4.8.2 模型的性能分析

另外，本文认为，“陈述”框架的语义角色的标注性能较低的主要原因有以下几个方面：

- (1) 语义角色的边界识别率低。本文将测试集中各个语义角色块的标注分为三

种情况:

情况 1: 语义角色块边界识别不正确。

情况 2: 语义角色块边界识别正确, 但是语义角色类型识别错误。

情况 3: 语义角色块的边界和语义角色类型全部识别正确

表 4.14 中详细列出了三种情况下, 各语义块的频次分布。从表中可以明显看出, 边界正确情况下, 角色类别的标注不正确的块很少, 绝大多数情况下是边界识别的错误, 名词性词元尤其明显。

表 4.14 “陈述”框架下测试集中语义角色块的错误情况统计

语义角色	动词模型				名词模型			
	总块数	情况 1	情况 2	情况 3	总块数	情况 1	情况 2	情况 3
medium	280	76	14	190	24	22	0	2
spkr	417	134	2	281	76	44	0	32
msg	647	189	0	458	4	3	1	0
核心语义角色	1344	399	16	929	104	69	1	34
time	133	66	4	63	13	9	1	3
manr	101	59	12	30	57	17	0	40
depic	25	18	2	5	0	0	0	0
cau	1	1	0	0	0	0	0	0
top	14	11	2	1	70	32	1	37
degr	1	1	0	0	6	0	6	0
add	138	39	7	92	0	0	0	0
非核心语义角色	413	195	27	191	146	58	8	80
总计	1757	594	43	1120	250	127	9	114

表 4.14 的统计结果让本文不得不思考是否应该把语义角色边界的识别作为一个单独的子任务进行研究。为此, 本文使用 CRF 模型进行了初步的边界识别实验。其中, 仍然使用正交表 $L_{32}(4^9 \times 2^4)$ 进行特征选择, 并从中挑选出最优的边界识别结果。边界识别的初步实验结果表明, 动词边界的准确率, 召回率, F1-值为 77.06%, 70.75%, 73.77%, 而边界识别与角色分类同时作的结果为 76.17%, 64.60, 69.91%, F1-值只相差 3.86%。名词边界的准确率, 召回率, F1 值为 70.17%, 50.80, 58.93%, 而边界识别与角色分类同时作的结果为 77.55%, 45.60%, 57.43%, F1-值只相差 1.5%。因此, 将标注步骤分为两步 (边界识别和类别识别) 对性能不会显著提高 (两步都用 CRF 模型时)。所以, 在语义角色边界识别和角色分类合起来做, 对 CRF 模型还是可以的。

(2) 词特征的影响。在特征集中去掉词特征, 动词标注的结果准确率, 召回率,

F1-值为 65.59%, 54.35%, 59.45%, 有大约 9.77% 的下降; 名词为 53.98%, 38.00%, 44.60%, 有大于 12.98% 的下降, 因此, 词特征是关键的特征。要使语义角色块标注正确, 首先使得组成其语义角色块的词的标注正确, 而从词的标注(而不是从语义角色块)情况看, 那些在测试集中出现而在训练集中未出现的词的标注正确率较其他词明显低。“陈述”框架下, 分别针对动词词元和名词词元进行统计, 得到如表 4.15 所示。

表 4.15 “陈述”框架语义角色中基于词的标注结果

动词	角色中, 在测试集中出现, 也在训练集中出现的词语数目	7764	标注正确的词数	6001	正确率	77.29%
			标注错误的词数	1763	错误率	22.71%
	角色中, 在测试集中出现, 在训练集中没有出现的词语数目	1979	标注正确的词数	1435	正确率	72.51%
			标注错误的词数	544	错误率	27.49%
名词	角色中, 在测试集中出现, 也在训练集中出现的词语数目	888	标注正确的词数	418	正确率	47.07%
			标注错误的词数	470	错误率	52.93%
	角色中, 在测试集中出现, 在训练集中没有出现的词语数目	280	标注正确的词数	89	正确率	31.79%
			标注错误的词数	191	错误率	68.21%

从表 4.15 中看出, 那些在测试集中出现而在训练集中未出现的词的标注正确率动词会下降约 5%, 名词会下降约 16%, 总体下降约 9%。因此, 测试集中出现而未在训练集中出现的词是性能下降的一个主要因素。

(3) 缺乏句法层面的特征。无论是英文还是中文的语义角色自动标注研究, 句法特征是主要的候选特征。Nianwen Xue(2008)的研究表明, 在人工标注好的完全句法分析树上进行语义角色的标注 F-值能达到 94.1%, 可与英文的结果相媲美。但是, 如果使用自动的句法分析的基础上, 再进行语义角色的标注却只有 71.9%(在已分好词、标好词性的测试集例句上)。但是, 目前汉语的句法分析器还很不成熟, 对开发领域的分析结果还达不到可用的地步。

4.8.3 “陈述”框架语义角色标注性能曲线

随机抽取动词“陈述”框架中动词词元的训练集语料库中例句数, 使其例句个数 n 与框架定义的元素个数 e 的比值 $r=n/e$, 分别满足 $r \leq 3$, $3 < r \leq 6$, $6 < r \leq 9$, $9 < r \leq 12$, $12 < r \leq 15$, $15 < r$ 六部分, 看其大小对语义角色标注性能(所有框架元素的 F 值)的影响, 如图 4.2 和图 4.3 所示。

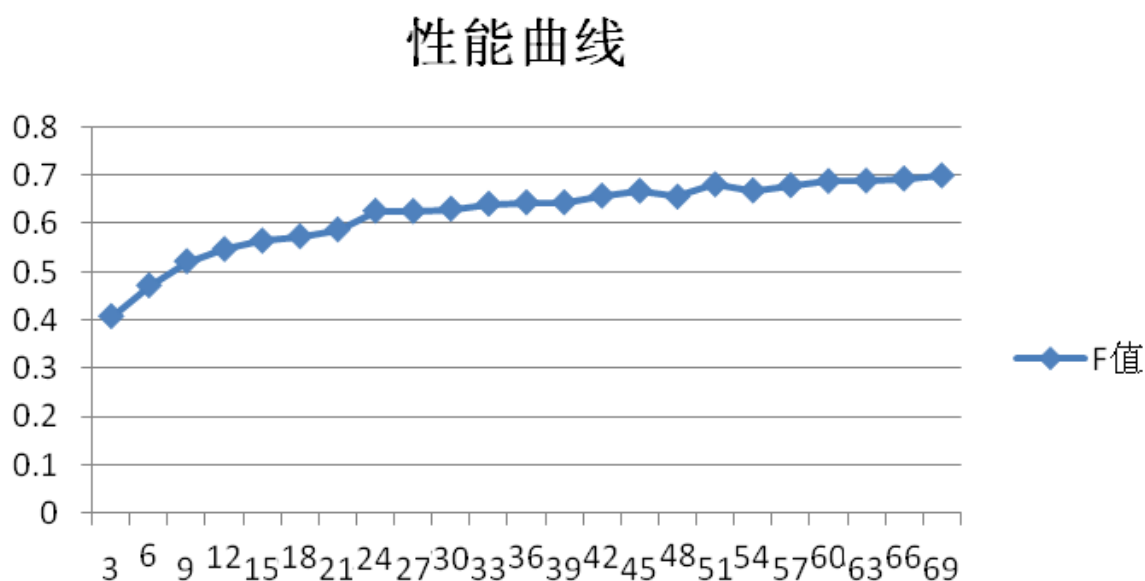


图 4.2 “陈述”框架下动词词元的例句个数与语义角色标注的性能曲线

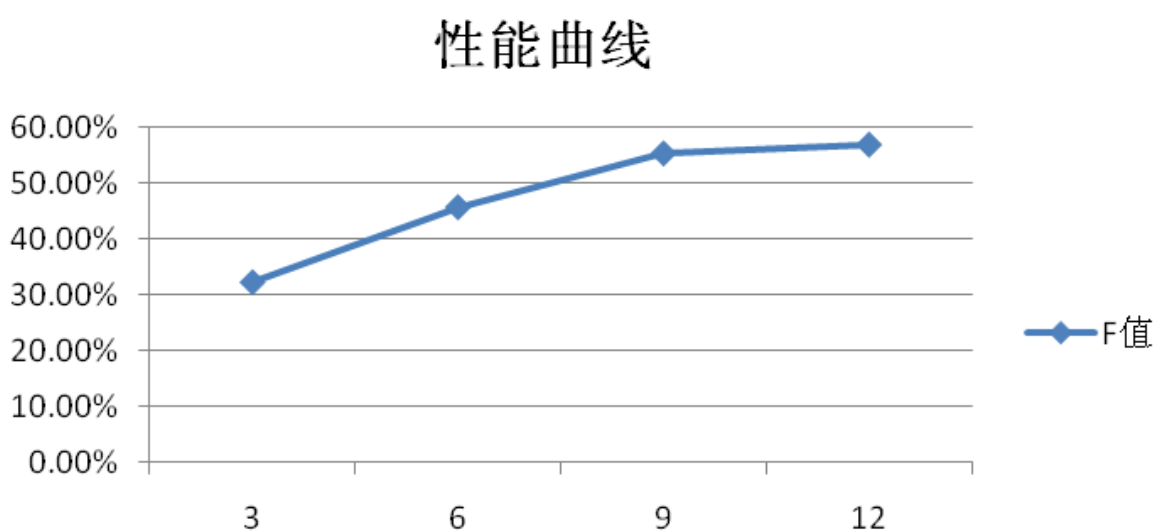


图 4.3 “陈述”框架下名词词元的例句个数与语义角色标注的性能曲线

从图中可以看出,随着例句个数 n 与框架定义的元素个数 e 的比值 $r=n/e$ 的增大,其标注的性能呈现上升趋势。换句话说,如果继续增加语料库规模,模型的性能还有增加的余地。

4.9 其它框架的实验结果

表 4.16 中给出了在所选的 25 个框架上的实验结果,表中分别给出了核心语义角色、非核心语义角色和所有语义角色的标注精度。对于所选择的 25 个框架,如果分词性框架的分布显著不同,则看作不同框架。对 15 个框架显著性检验的结果为:

只有“陈述”和“记忆”的不同词性词元有显著区别，但“记忆”框架的其它词性词元的例句较少（20句），无法单独训练模型，故不再区分。其它框架在目前的例句规模下都不显著，也即不需要分词性。对每个框架的例句库，按 5:5 拆分为训练集、测试集，分别建模、训练和测试。

表 4.16 选出的 25 个框架在测试集上的标注结果

框架名	模板	核心语义角色			非核心语义角色			所有语义角色		
		P	R	F	P	R	F	P	R	F
联想	11	47.50%	25.85%	33.48%	0.00%	0.00%	0.00%	47.50%	25.85%	33.48%
范畴化	1	73.68%	31.82%	44.44%	60.00%	11.54%	19.35%	72.58%	28.48%	40.91%
查看	16	65.75%	30.57%	41.74%	68.29%	37.33%	48.28%	67.10%	33.88%	45.02%
发明	6	68.35%	50.26%	57.93%	83.33%	8.33%	15.15%	68.97%	40.16%	50.76%
思想	4	68.42%	41.94%	52.00%	0.00%	0.00%	0.00%	68.42%	41.94%	52.00%
因果	12	77.14%	45.38%	57.14%	100.00%	6.67%	12.50%	77.46%	41.04%	53.66%
非自主感知	20	71.90%	56.31%	63.16%	66.67%	21.28%	32.26%	71.32%	48.14%	57.48%
陈述_n	21	79.07%	32.69%	46.26%	77.67%	54.79%	64.26%	78.08%	45.60%	57.58%
感受	12	53.41%	46.08%	49.47%	73.04%	69.42%	71.19%	61.17%	54.77%	57.79%
自主感知	6	72.03%	59.86%	65.38%	56.10%	17.04%	26.14%	70.38%	49.55%	58.16%
宗教信仰	6	80.00%	50.00%	61.54%	0.00%	0.00%	0.00%	80.00%	48.28%	60.22%
外观	10	73.33%	54.32%	62.41%	100.00%	28.57%	44.44%	74.19%	52.27%	61.33%
思考	10	77.70%	57.75%	66.26%	72.73%	20.00%	31.37%	77.33%	51.10%	61.54%
适宜性	11	69.23%	57.69%	62.94%	0.00%	0.00%	0.00%	69.23%	55.56%	61.64%
相信	1	80.81%	60.61%	69.26%	100.00%	26.15%	41.46%	83.62%	49.24%	61.98%
获知	16	71.63%	57.89%	64.03%	84.62%	39.29%	53.66%	73.03%	54.66%	62.52%
记忆	6	79.69%	59.30%	68.00%	83.33%	29.41%	43.48%	80.18%	51.90%	63.01%
证明	20	67.61%	63.16%	65.31%	100.00%	37.50%	54.55%	70.13%	58.70%	63.91%
计划	1	77.42%	57.14%	65.75%	70.00%	63.64%	66.67%	76.39%	57.89%	65.87%
鲜明性	29	82.86%	64.80%	72.73%	100.00%	3.45%	6.67%	82.98%	56.25%	67.05%
拥有	1	79.20%	60.00%	68.28%	0.00%	0.00%	0.00%	79.20%	59.28%	67.81%
陈述_v	1	75.47%	69.12%	72.16%	77.02%	46.25%	57.79%	75.73%	63.75%	69.22%
代表	12	79.41%	64.29%	71.05%	100.00%	37.50%	54.55%	80.28%	61.96%	69.94%
知觉特征	1	83.20%	61.18%	70.51%	85.94%	66.27%	74.83%	84.13%	62.85%	71.95%
属于某类	20	88.33%	71.62%	79.10%	0.00%	0.00%	0.00%	86.89%	69.74%	77.37%
包含	1	90.55%	88.46%	89.49%	0.00%	0.00%	0.00%	90.55%	88.46%	89.49%
总计		74.35%	58.28%	65.34%	75.94%	38.81%	51.37%	74.61%	53.77%	62.50%

表 4.16 中最后一行的总计结果是在 25 个框架的测试集上的微平均的标注准确率、召回率和 F1-值，而不是 25 个框架的宏平均结果。从表 4.16 中，本文可以得到如下几点结论：

(1)所有框架的自动标注结果都是准确率显著大于相应的召回率，召回率是影响

我们系统性能的主要指标。从“陈述”框架的错误分析，以及我们在若干框架上的初步试验得出，整个系统召回率低的主要原因是在每个框架的角色边界识别时的召回率很低。因此，提高角色边界识别的召回率应当是主攻方向。

(2)本文所选的 25 个框架很多都来自框架网的认知领域（如“联想”和“思想”），而认知域的框架即使在手工标注时都很难把握，因此其自动标注有一定的难度，具有挑战性。

(3)在 25 个框架上总体的标注 F 值可以达到 62.50%。其中，核心语义角色的 F 值为 65.34%，要明显高于非核心语义角色的 F 值 51.37%。初步分析，这主要是因为核心语义角色的数目较多，训练比较充分。

4.10 本章小结

本章将汉语框架语义角色标注问题看作是词层面的序列标注任务，并使用条件随机场模型对序列进行标注。为了权衡训练数据稀疏和语义角色搭配模型的分布两方面的因素，本章提出了使用卡方检验来检验同一个框架下不同词性的语义角色分布情况。在此基础上，将一个框架下的语义角色标注模型按照词元词性显著程度分裂成几个独立的语义角色标注模型进行求解。

为了避免传统经验选取的盲目性，本章使用正交表对最优的特征模板进行了选择。实验结果清晰地表明正交表特征选择方法找到的最优特征模板优于人工主观选取的模板。本章进一步使用正交表的极差分析方法对模型中各类特征的重要程度进行了分析，并按照极差的大小给出了一个特征的重要程度的排列顺序。

为了证明方法的有效性和通用性，本文针对人工选择的 25 个框架进行了实验。最终的实验结果表明，25 个框架的总体语义角色标注 F 值可以达到 60%以上，这是目前汉语框架语义角色标注性能的一个初步结果。

第5章 融合汉语基本块信息的汉语框架语义角色标注模型

本章将基本块信息引入到汉语框架语义角色标注模型中。原有 CFN 例句库没有正确的基本块标注信息，因此，本章使用的基本块信息均为汉语基本块分析器自动产生的。本章中，本文首先分析了汉语句子的基本块和语义角色之间的关系，并提出了两种汉语框架语义角色标注模型：“基于词的标注模型”和“基于基本块的标注模型”。针对两种模型，本文分别从汉语基本块信息中提取出相应的特征。然后，使用正交表设计了多组特征组合实验。最后，得到了相应的实验结果。

5.1 汉语基本块描述体系及基本块自动标注性能

使用 Abney 的块理论，结合汉语的特点，周强定义了汉语基本块的描述体系^[75]，即：基本块 = 基本拓扑结构 + 句法形式描述 + 语义内容描述，并给出了相应的基本块标记集合。该体系认为，汉语基本块的主要特点是块内部的各个词语按照一定的句法关系组合在以一词为中心的结构上，并可以通过这个中心词来体现整个基本块的外部功能。在构建基本块语料时，根据块内词语数目的不同，将基本块分为单词块和多词块，并将多词块归纳到三种基本的拓扑结构中，即：左角中心结构（LCC）、右角中心结构（RCC）和链式关联结构（CHC）。

每个基本块使用两种标记的组合：句法标记和关系标记，来描述它们的结构组合和语法关系。表 5.1 列出本文目前所用的句法标记集合和关系标记集合。

表 5.1 中心聚合块的句法标记和关系标记描述集

句法标记	内容描述	关系标记	内容描述
np	名词块	ZX	右角中心结构
mp	数量块	LN	链式关联结构
sp	空间块	LH	并列关系 CHC
tp	时间块	PO	述宾关系 LCC
vp	动词块	SB	述补关系 LCC
ap	形容词块	AD	附加关系 LCC
dp	副词块	JB	介宾关系 LCC
pp	介词块	SG	单词语块

针对汉语基本块，周强使用规则的方法开发了汉语基本块自动分析器^[76]。为了适应对真实文本中各种复杂语言现象的分析需要，汉语基本块自动分析器的自动分析结果中还增加了以下两个关系标记：RL 和 AM，其具体描述如下：

RL：复杂关系描述，在现有知识条件下很难确定基本块的内部结构符合目前定

义的三种拓扑结构中的哪一种时，使用此标记表示；

AM：歧义区间描述，表示该区间中的若干词语可以形成不同的结构组合关系，根据现有知识很难进行自动排歧，只能保留多个结构组合留待后续系统进行选择处理。

分析器的输入是经过切分和词性标注的汉语句子，它的输出则标注出了每个基本块的边界以及语法标记和关系标记。下面是基本块分析器一个输入输出的例子。

分析器输入：英 jn 方面 n 作为 v 报复 v 措施 n 也 d 宣布 v 4 m 名 q 俄罗斯 nsy 大使馆 n 的 u 外交官 n 为 v 不 d 受 v 欢迎 v 的 u 人 n 。 w

分析器输出：[np-AM 英/jn 方面/n] [vp-SG 作为/v] [vp-PO 报复/v 措施/n] [dp-SG 也/d] [vp-SG 宣布/v] [mp-ZX 4 /m 名/q] [np-ZX 俄罗斯/nsy 大使馆/n] 的/u [np-SG 外交官/n] [vp-AM 为/v 不/d 受/v 欢迎/v] 的/u [np-SG 人/n] 。 /w

由上面的例子可以看出，基本块分析器是把一个输入的句子中的连接紧密的词按从底向上组合的方法生成一个个基本块。表 5.2 为基本块分析器的实验结果，可以看出基本块分析器整体的性能还是比较理想的。

表 5.2 基本块分析器整体实验结果

句法 标记	封闭测试			开放测试		
	正确率	召回率	F-M	正确率	召回率	F-M
np	90.32%	88.74%	89.53%	88.33%	86.62%	87.46%
vp	91.13%	92.36%	91.74%	89.44%	90.58%	90.00%
mp	95.00%	96.50%	95.75%	92.18%	94.72%	93.43%
ap	93.55%	93.45%	93.50%	94.28%	95.65%	94.96%
tp	91.49%	91.84%	91.67%	89.97%	89.82%	89.89%
sp	81.88%	88.08%	84.87%	80.12%	88.97%	84.31%
合计	91.04%	91.39%	91.21%	89.33%	89.61%	89.47%

5.2 基本块和汉语框架语义角色的关系

目前，语言学家普遍认为，语义角色块和完全句法分析树中的一个成分相对应。因此，在进行英语的语义角色自动标注时，很多学者们使用完全句法分析树作为输入，通过模型来判断分析树上的哪些成分是语义角色，哪些不是。由于英语中自动完全句法分析器精度较高，这种做法是可行的。

但是，汉语的完全句法分析性能还不尽人意。本文不可能效仿英文的做法来进行汉语的语义角色自动标注。但是，汉语中浅层组块自动分析的精度还是不错的。

周强提出的基本块描述体系便属于汉语组块体系中的一种。基本块描述的是汉语句子里联系紧密的实义词之间的关系。周强认为汉语句法分析树上的每个成分都是由基本块进一步聚合形成的。

本文认为,对于汉语句子中的 CFN 语义角色块是由若干个基本块组成的。由于 CFN 句子库中没有正确的基本块标记,本文只能使用自动分析得到的汉语基本块信息来粗略地研究它们之间的关系。本文对 CFN 句子库整体进行了汉语基本块自动分析(词性映射表见附录 A),清华大学的词性标注规范请参见文献[77]。在此基础上,统计了“陈述”框架中动词词元和名词词元的句子中语义角色和基本块的对应情况,见表 5.3 和表 5.4 中“处理前”的数据。通过分析,本文制订了如下两条后处理规则对自动分析得到的基本块信息进行校正。

(1) 将关系标记为“AM”和“RL”的基本块打散成单词块或块外词。

(2) 如果一个基本块的词语序列中含有目标词元,并且该基本块由多个词语组成,则将该基本块打散成单词块或块外词。

基本块的具体的打散规则如下:

(1) 如果词语的词性标记首字母属于: k-c-e-l-p-u-y-w,则本文将其确定为块外词。否则,该词语为实义词,本文将其上升为单词块。

(2) 对其他实义词,按照下列规则提升不同词类标记为相应基本块句法标记:
aN --> np; aD --> dp; rD --> dp; rV --> vp; vN --> np; qT --> tp; n*, r*, j, x, g --> np; f, s --> sp; b, a, z, o --> ap; d --> dp; v --> vp; m --> mbar; q --> mp; t --> tp; iV --> vp; i, iN --> np;并将该实义词的关系标记全部提升为 SG;

举例说明,使用启发式规则对 5.1 节中给出的基本块分析结果处理后,得到的基本块标注序列如下:

经过校正的基本块信息: [np-SG 英/jn] [np-SG 方面/n] [vp-SG 作为/v] [vp-PO 报复/v 措施/n] [dp-SG 也/d] [vp-SG 宣布/v] [mp-ZX 4/m 名/q] [np-ZX 俄罗斯/nsy 大使馆/n] 的/u [np-SG 外交官/n] [vp-SG 为/v] [dp-SG 不/d] [vp-SG 受/v] [vp-SG 欢迎/v] 的/u [np-SG 人/n] 。/w

本文使用这种启发式的方法对 CFN 句子库的自动标注的基本块信息进行了后处理,并再次统计了“陈述”框架中动词词元和名词词元的句子中语义角色和基本块的对应情况,具体见表 5.3 和表 5.4 中“处理后”的数据。从中可以看出,经过处理后,绝大多数的基本块与 CFN 语义角色块的边界是可以对齐的。这些数据进一步说明了“基本块是 CFN 语义角色的组成单位”这一语言事实。不过,从表中可以发

现，仍然有 2%-3%的语义角色块的边界与基本块的边界无法对齐。导致这种状况的主要原因是基本块信息是自动分析得到的；另一方面的原因是标注体系本身存在着不一致现象。

表 5.3 “陈述”框架动词词元例句中语义角色块与基本块的边界对应情况

语义角色	处理前				处理后			
	原有	可对齐	对不齐	错误率	原有	可对齐	对不齐	错误率
degr	2	1	1	50.00%	2	2	0	0.00%
spkr	832	780	52	6.25%	832	795	37	4.45%
msg	1302	1157	145	11.14%	1302	1284	18	1.38%
time	273	256	17	6.23%	273	271	2	0.73%
manr	205	175	30	14.63%	205	194	11	5.37%
depic	57	49	8	14.04%	57	55	2	3.51%
cau	3	3	0	0.00%	3	3	0	0.00%
top	42	41	1	2.38%	42	42	0	0.00%
add	264	178	86	32.58%	264	257	7	2.65%
medium	540	524	16	2.96%	540	532	8	1.48%
合计	3520	3164	356	10.11%	3520	3435	85	2.41%

表 5.4 “陈述”框架名词词元例句中语义角色块与基本块的边界对应情况

语义角色	处理前				处理后			
	原有	可对齐	对不齐	错误率	原有	可对齐	对不齐	错误率
degr	6	4	2	33.33%	6	6	0	0.00%
spkr	157	141	16	10.19%	157	148	9	5.73%
msg	9	8	1	11.11%	9	8	1	11.11%
time	32	29	3	9.38%	32	31	1	3.13%
manr	117	60	57	48.72%	117	115	2	1.71%
depic	1	0	1	100.00%	1	1	0	0.00%
cau	0	0	0	0.00%	0	0	0	0.00%
top	144	136	8	5.56%	144	144	0	0.00%
add	0	0	0	0.00%	0	0	0	0.00%
medium	47	42	5	10.64%	47	44	3	6.38%
合计	513	420	93	18.13%	513	497	16	3.12%

5.3 融合基本块信息的汉语框架语义角色标注模型

由于基本块是 CFN 语义角色的组成单位，因此，将基本块信息引入到语义角色标注模型后，可以尝试分别从“词层面”和“基本块层面”建立相应的语义角色标注模型。具体如下：

在词层面，模型的建立方法和 4.1.1 小节中给出的一样，以词语作为标注单位，使用 IOB2 策略来刻画每个词语与 CFN 语义角色块的关系。本文将该模型称为“基于词的模型”。在本章中，该模型除了使用词层面的特征外，还使用了基本块特征以及其它一些新的特征。具体的特征描述及选取方法见 5.4.1 小节。

以基本块作为 CFN 语义角色标注模型的标注单位，模型的形式化方法如下：

针对一个 CFN 框架中的所有的框架元素 X ，本文设立相应的标注集合 $\{B-X, I-X, O\}$ ，并对每个基本块以及基本块的块外词赋予一个标记，来实现 CFN 语义角色块的自动标注。其中，前 2 个标记是由两部分组成。第一部分代表基本块或块外词在语义角色块中的位置，如果基本块或块外词在语义角色块的开始位置则使用 B 进行标记，如果在内部位置则使用 I 进行标记。否则，本文统一使用 O 进行标记。本文将这种模型称为“基于基本块的模型”。

例句 1 的语义角色块序列在基本块层面的 IOB2 标注如下（这里，本文使用基本块的中心词表示整个基本块）：

英|B-spkr 方面|I-spkr 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

上面的例子中，本文发现使用基本块信息对句子进行压缩后，句子长度只是略有变短，这主要是因为该句子中的单词块(SG)比较多。“基于基本块的模型”中所使用到的特征集合及具体的描述在 5.5.2 小节给出。

5.4 基于词的模型

5.4.1 模型的特征描述及选取

使用词语作为基本块的标注单位，本文分别从词层面和基本块层面进行了特征选取和优化，具体选取的特征描述如下：

1 词层面特征

词特征：指当前所标注的词条。

词性特征：指当前标注的词条的词性。

位置特征：指当前标注的词语与目标词之间的位置关系；

目标词特征：指当前标注的词语所在句子的目标词元。

命名实体的特征：使用一个标注出当前词是否是命名实体。具体的命名实体需要通过词性的映射来决定。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、nsy=外国地名、ni=机构专名}。TIME 包括{nt=时间名词}。

是否为简单句 (二元特征): 使用"是"和"否"来表示当前句子是否为简单句。

2 基本块层面特征

句法标记特征: 指当前标注的词所具有的句法标记。

关系标记特征: 指当前标注的词所具有的关系标记。

块中心词特征: 指当前标注的词所在基本块及相邻块的中心词词条。

块中心词的词性特征: 指当前标注的词所在基本块及相邻块的中心词的词性。

线性句法路径特征: 指由当前所标注的词语和谓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特征。

浅层子范畴特征: 描述谓词词元的基本块周围的基本块组成的序列,该特征开一个窗口。

从上述的特征集合中可以看出,"词"、"词性"、"位置"和"目标词"特征在 BASELINE 系统中出现过。句法标记特征和关系标记特征分别使用基本块的句法标记和关系标记作为特征,并使用 IOB2 策略来刻画。块中心词特征指的是基本块的当前中心词以及左右相邻块的中心词。

表 5.5 中给出了基于词的基本块模型中针对上述三大类特征的候选特征类型和相应的窗口大小。其中,本文将词特征、词性特征、位置特征、句法标记特征、关系标记特征、中心词特征和中心词词性特征七类特征看做是实验中必须包含的特征,并将它们的最小窗口设置为[0,0]。一个句子中的每个词的目标词特征都是相同的,因此,本文将它们作为必选特征,并将其窗口仅设定为[0,0]。

除了上述的必选特征之外,本文还将每一种必选特征的组合特征和不同特征之间的搭配特征加入到候选特征集合中。对于这些特征,本文将它们的最低水平设置为“-”,即,不选取该特征。之所以这样设置是因为:第一,组合和搭配特征的稀疏程度较严重,特征空间较大,强行加入可能会导致模型过拟合程度增加。第二,这些特征是由必选特征组合或搭配形成的,在引入新的特征信息的同时,可能也引入了一些噪声特征。

此外,浅层句法路径特征、命名实体特征、框架子范畴特征和是否简单句特征,刻画了当前标注单位的一些结构特征和语义信息,但是使用这些特征的上下文组合和搭配特征意义不大。因此,本文仅考虑了“选取”和“不选取”(窗口为[0,0])两

种水平。为了便于使用正交表进行实验,表 5.5 中将这四类特征的水平 3 设置为拟水平,即,水平 3 的窗口和水平 2 相同。

表 5.5 基于词的标注模型的候选特征及其窗口大小

编号	特征	水平 1	水平 2	水平 3
1	词	[0, 0]	[-1, 1]	[-2, 2]
2	词二元组	-	[-1, 1]	[-2, 2]
3	词性	[0, 0]	[-1, 1]	[-2, 2]
4	词性二元组	-	[-1, 1]	[-2, 2]
5	位置	[0, 0]	[-1, 1]	[-2, 2]
6	位置二元组	-	[-1, 1]	[-2, 2]
7	位置三元组	-	[-1, 1]	[-2, 2]
8	词与词性的搭配	-	[0, 0]	[-1, 1]
9	词与位置的搭配	-	[0, 0]	[-1, 1]
10	词性与位置的搭配	-	[0, 0]	[-1, 1]
11	句法标记	[0, 0]	[-1, 1]	[-2, 2]
12	句法标记二元组	-	[-1, 1]	[-2, 2]
13	句法标记三元组	-	[-1, 1]	[-2, 2]
14	关系标记	[0, 0]	[-1, 1]	[-2, 2]
15	关系标记二元组	-	[-1, 1]	[-2, 2]
16	中心词	[0, 0]	[-1, 1]	[-2, 2]
17	中心词二元组	-	[-1, 1]	[-2, 2]
18	中心词性	[0, 0]	[-1, 1]	[-2, 2]
19	中心词性二元组	-	[-1, 1]	[-2, 2]
20	句法与中心词搭配	-	[0, 0]	[-1, 1]
21	关系与中心词搭配	-	[0, 0]	[-1, 1]
22	浅层句法路径	-	[0, 0]	[0, 0]
23	命名实体	-	[0, 0]	[0, 0]
24	浅层子范畴	-	[0, 0]	[0, 0]
25	是否简单句	-	[0, 0]	[0, 0]
26	目标词(必选)	[0,0]		

根据表 5.5 中给出的候选特征集合及相应的窗口大小,本文决定采用具有 54 组实验的 PB 表 $PB_{54}(3^{25} \times 2^1)$ 进行实验。实验中,本文中仅使用了 PB 表中前 25 个 3 水平列。PB 表是一种非正规正交表,具体理论可以参照文献[78]。

5.4.2 模型的实验结果

本文针对所选择的 25 个框架进行了实验。由于篇幅有限,本文仅给出了“陈述”框架下动词模型和名词模型的试验结果。其中,本文选择所有框架语义角色的 F 值最高时所对应的模板作为最优的试验模板。

1. “陈述”框架动词模型的试验结果

表 5.6 中给出了“陈述”框架动词模型的 54 组实验结果。从中可以看出,模板 38 的总体 F 值是最高的,达到了 70.56%。与仅使用词层面信息的最好总体 F 值 69.22% 相比,上升了 1.3% 左右。并且,核心框架元素和非核心框架元素的总体 F 值也相应的有所上升。

表 5.6 “陈述”框架动词模型的 54 组实验结果

模板 编号	非法 序列	特征数	训练时间	核心框架语义角色			非核心框架语义角色			所有框架语义角色		
				P	R	F	P	R	F	P	R	F
38	5	3105340	536.70s	78.83%	68.97%	73.57%	80.93%	46.25%	58.86%	79.18%	63.63%	70.56%
48	6	2712860	541.92s	77.43%	69.94%	73.49%	81.12%	45.76%	58.51%	78.02%	64.26%	70.47%
31	4	2360940	588.25s	77.69%	69.20%	73.20%	81.43%	46.73%	59.38%	78.31%	63.92%	70.39%
42	4	3153760	567.09s	77.84%	69.79%	73.60%	79.57%	45.28%	57.72%	78.12%	64.03%	70.38%
26	4	3089400	455.56s	78.88%	68.90%	73.55%	80.53%	44.07%	56.96%	79.14%	63.06%	70.19%
11	7	2329200	528.78s	76.74%	68.97%	72.65%	80.97%	48.43%	60.61%	77.46%	64.14%	70.17%
10	2	2322240	587.66s	77.28%	69.35%	73.10%	79.10%	46.73%	58.75%	77.59%	64.03%	70.16%
41	5	2813520	432.70s	77.59%	69.57%	73.36%	78.57%	45.28%	57.45%	77.75%	63.86%	70.12%
15	4	1887720	539.05s	77.62%	69.42%	73.29%	77.69%	45.52%	57.40%	77.63%	63.80%	70.04%
25	5	3456160	467.77s	77.65%	68.75%	72.93%	80.79%	44.79%	57.63%	78.15%	63.12%	69.84%
22	4	2691300	504.30s	77.46%	68.01%	72.42%	80.83%	46.97%	59.42%	78.03%	63.06%	69.75%
2	7	2588800	539.89s	77.01%	68.53%	72.52%	81.20%	46.00%	58.73%	77.69%	63.23%	69.72%
32	6	2135980	484.20s	77.32%	68.23%	72.49%	80.08%	46.73%	59.02%	77.79%	63.18%	69.72%
37	2	2968960	498.48s	78.47%	68.60%	73.20%	78.85%	43.34%	55.94%	78.53%	62.66%	69.71%
28	4	2133780	467.23s	77.86%	68.82%	73.06%	77.87%	44.31%	56.48%	77.86%	63.06%	69.69%
9	3	1796600	585.63s	76.59%	69.12%	72.66%	78.99%	45.52%	57.76%	76.98%	63.57%	69.64%
1	5	2483140	516.67s	76.92%	68.68%	72.56%	79.75%	45.76%	58.15%	77.38%	63.29%	69.63%
33	5	2648360	492.92s	77.53%	68.53%	72.75%	79.22%	44.31%	56.83%	77.80%	62.83%	69.52%
19	4	2244040	536.73s	77.01%	68.30%	72.40%	79.50%	46.00%	58.28%	77.43%	63.06%	69.51%
6	3	2378420	546.52s	77.98%	68.23%	72.78%	78.54%	44.31%	56.66%	78.07%	62.61%	69.49%
36	5	3881420	591.33s	77.50%	67.93%	72.40%	80.52%	45.04%	57.76%	78.00%	62.55%	69.43%
8	5	1406000	614.66s	75.94%	69.05%	72.33%	75.89%	46.49%	57.66%	75.93%	63.75%	69.31%
14	3	1900100	616.95s	76.37%	69.27%	72.65%	76.79%	44.07%	56.00%	76.44%	63.35%	69.28%
39	2	2587920	584.97s	77.51%	68.23%	72.58%	78.70%	43.83%	56.30%	77.71%	62.49%	69.27%
35	4	2152160	642.73s	77.19%	68.23%	72.43%	78.54%	44.31%	56.66%	77.41%	62.61%	69.23%
7	6	1569900	655.81s	76.03%	68.90%	72.29%	75.49%	46.25%	57.36%	75.93%	63.57%	69.21%
3	3	2132160	521.67s	76.99%	68.23%	72.35%	77.97%	44.55%	56.70%	77.15%	62.66%	69.16%
30	3	3315780	528.33s	78.07%	68.08%	72.73%	78.48%	42.37%	55.03%	78.14%	62.04%	69.16%
16	3	2195740	583.61s	76.46%	68.15%	72.07%	78.66%	45.52%	57.67%	76.83%	62.83%	69.13%
29	4	2639960	554.47s	77.43%	67.63%	72.20%	79.06%	44.79%	57.19%	77.70%	62.27%	69.13%
52	4	2572500	517.16s	77.18%	67.93%	72.26%	79.82%	44.07%	56.79%	77.60%	62.32%	69.13%
34	7	2934080	539.45s	77.07%	68.01%	72.25%	80.45%	42.86%	55.92%	77.60%	62.09%	68.99%
23	5	2999180	627.91s	77.22%	67.86%	72.24%	78.95%	43.58%	56.16%	77.50%	62.15%	68.98%
40	3	3193160	527.47s	77.04%	67.41%	71.90%	80.97%	44.31%	57.28%	77.67%	61.98%	68.95%

4	3	1795360	500.94s	76.38%	67.86%	71.87%	78.57%	45.28%	57.45%	76.75%	62.55%	68.92%
53	6	2693200	376.78s	77.65%	67.49%	72.21%	76.92%	43.58%	55.64%	77.53%	61.87%	68.82%
46	4	2859540	474.16s	76.92%	67.71%	72.02%	79.82%	43.10%	55.97%	77.38%	61.92%	68.80%
13	7	1328920	564.70s	75.76%	68.38%	71.88%	74.90%	45.52%	56.63%	75.61%	63.01%	68.74%
24	4	1597380	541.64s	76.36%	68.01%	71.94%	76.79%	44.07%	56.00%	76.43%	62.38%	68.69%
5	6	2436340	486.56s	76.77%	67.11%	71.62%	79.91%	44.31%	57.01%	77.28%	61.75%	68.65%
20	5	3310780	552.80s	76.72%	67.19%	71.64%	80.18%	44.07%	56.88%	77.28%	61.75%	68.65%
17	6	3351140	542.53s	77.32%	66.96%	71.77%	80.27%	43.34%	56.29%	77.79%	61.41%	68.64%
21	5	2018340	538.36s	76.28%	67.49%	71.61%	76.95%	45.28%	57.01%	76.40%	62.27%	68.61%
50	6	3446580	440.72s	77.39%	66.96%	71.80%	79.46%	43.10%	55.89%	77.72%	61.35%	68.58%
12	6	1866580	528.23s	76.79%	67.19%	71.67%	78.51%	43.34%	55.85%	77.07%	61.58%	68.46%
47	3	4040800	417.00s	77.55%	66.07%	71.35%	81.08%	43.58%	56.69%	78.13%	60.79%	68.37%
45	5	2148640	588.00s	76.80%	66.74%	71.42%	78.35%	43.83%	56.21%	77.06%	61.35%	68.31%
51	3	3355780	443.80s	76.81%	67.04%	71.59%	78.32%	42.86%	55.40%	77.06%	61.35%	68.31%
43	6	2658920	468.58s	76.66%	66.00%	70.93%	79.06%	44.79%	57.19%	77.07%	61.01%	68.11%
49	5	3626720	450.95s	77.90%	65.85%	71.37%	81.64%	40.92%	54.52%	78.48%	59.99%	68.00%
44	6	3373060	488.73s	77.16%	66.37%	71.36%	78.80%	41.40%	54.29%	77.42%	60.50%	67.92%
27	4	2402880	469.06s	75.86%	65.92%	70.54%	78.60%	43.58%	56.07%	76.31%	60.67%	67.60%
18	3	1884560	577.27s	75.02%	65.48%	69.92%	77.31%	44.55%	56.53%	75.41%	60.56%	67.17%
0	11	190640	448.52s	61.99%	50.60%	55.71%	61.90%	34.62%	44.41%	61.97%	46.84%	53.35%

模板 38 对应的具体的特征组合方式见表 5.7。显而易见，正交表综合考虑了词层面和基本块层面的所有特征。最优的实验结果的细节在表 5.8 中给出。其中，核心框架元素的总体识别 F 值可以达到 73.57%，但是，非核心框架元素的总体识别 F 值仅为 58.86%。两者相差 14.71%。这主要是由于训练集中，核心框架元素的块数明显多余非核心框架元素的块数（近 3 倍），使得模型参数产生了严重的偏执现象。

表 5.7 模板 38 对应的特征组合方式

编号	特征	选取方式	编号	特征	选取方式
1	词	[-2, 2]	14	关系标记	[-1, 1]
2	词二元组	[-2, 2]	15	关系标记二元组	[-1, 1]
3	词性	[0, 0]	16	中心词	[0, 0]
4	词性二元组	[-2, 2]	17	中心词二元组	-
5	位置	[-2, 2]	18	中心词性	[0, 0]
6	位置二元组	[-1, 1]	19	中心词性二元组	[-2, 2]
7	位置三元组	-	20	句法与中心词搭配	[0, 0]
8	词与词性的搭配	-	21	关系与中心词搭配	[-1, 1]
9	词与位置的搭配	[-1, 1]	22	浅层句法路径	[0, 0]
10	词性与位置的搭配	[0, 0]	23	命名实体	[0, 0]
11	句法标记	[-1, 1]	24	浅层子范畴	-
12	句法标记二元组	-	25	是否简单句	[0, 0]
13	句法标记三元组	-	26	目标词	[0, 0]

表 5.8 模板 38 对应的“陈述”框架动词模型的最优实验结果

名字	原来	预测	正确	P	R	F
medium	280	207	187	90.34%	66.79%	76.80%
msg	647	617	458	74.23%	70.79%	72.47%
spkr	417	352	282	80.11%	67.63%	73.34%
核心元素	1344	1176	927	78.83%	68.97%	73.57%
depic	25	14	5	35.71%	20.00%	25.64%
degr	1	0	0	0.00%	0.00%	0.00%
manr	101	41	32	78.05%	31.68%	45.07%
add	138	102	95	93.14%	68.84%	79.17%
top	14	3	0	0.00%	0.00%	0.00%
time	133	76	59	77.63%	44.36%	56.46%
cau	1	0	0	0.00%	0.00%	0.00%
非核心元素	413	236	191	80.93%	46.25%	58.86%
合计	1757	1412	1118	79.18%	63.63%	70.56%

本文将以词语为单位，对融合基本块前与后的两个最优模型的各语义角色的 F 值（表 5.8 和表 4.7）进行了对比，得到了图 5.1 所示的结果。从图 5.1 中可见，加入基本块特征后，所有的核心框架元素及部分非核心框架元素的 F 值均有所提升，并且都在 70% 以上，但是，time 和 top 等非核心框架元素的 F 值却有所下降。

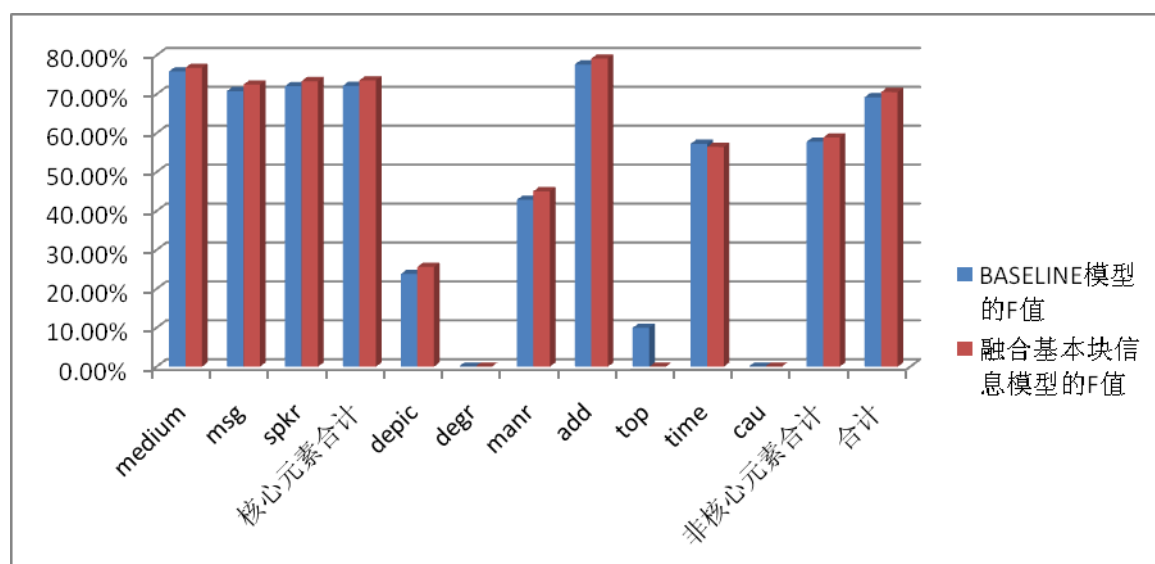


图 5.1 加入基本块信息前后动词模型各语义角色的 F 的对比

2. “陈述”框架名词模型的试验结果

表 5.9 给出了“陈述”框架中名词模型的 54 组实验结果。由于“陈述”框架下名词词元所拥有的例句较少，因此，模型的训练时间基本上仅维持在 30 秒左右，并

且实验结果也相对不太理想。从表 5.9 中可见, 模板 16 的实验结果最好, 总体 F 值仅达到了 54.45%, 较 BASELINE 模型中的最优 F 值 57.58%, 下降了 3% 左右。

表 5.9 “陈述” 框架名词模型的 54 组实验结果

模板 编号	非法 序列	特征数	训练时间	核心框架语义角色			非核心框架语义角色			所有框架语义角色		
				P	R	F	P	R	F	P	R	F
16	1	422128	29.23s	78.38%	27.88%	41.13%	73.58%	53.42%	61.90%	74.83%	42.80%	54.45%
2	2	482678	30.25s	71.43%	28.85%	41.10%	74.04%	52.74%	61.60%	73.29%	42.80%	54.04%
18	1	342692	25.77s	76.19%	30.77%	43.84%	70.37%	52.05%	59.84%	72.00%	43.20%	54.00%
38	0	568890	33.25s	81.58%	29.81%	43.66%	71.43%	51.37%	59.76%	74.13%	42.40%	53.94%
8	0	303506	28.66s	80.56%	27.88%	41.43%	70.27%	53.42%	60.70%	72.79%	42.80%	53.90%
25	2	630938	27.23s	78.38%	27.88%	41.13%	76.29%	50.68%	60.91%	76.87%	41.20%	53.65%
11	3	452186	30.27s	72.50%	27.88%	40.28%	71.56%	53.42%	61.18%	71.81%	42.80%	53.63%
17	1	605892	28.58s	78.38%	27.88%	41.13%	75.51%	50.68%	60.66%	76.30%	41.20%	53.51%
19	2	425628	29.80s	80.56%	27.88%	41.43%	71.03%	52.05%	60.08%	73.43%	42.00%	53.44%
9	2	355600	30.06s	73.17%	28.85%	41.38%	71.03%	52.05%	60.08%	71.62%	42.40%	53.27%
41	0	515424	30.61s	81.08%	28.85%	42.55%	71.15%	50.68%	59.20%	73.76%	41.60%	53.20%
32	1	412972	28.33s	80.56%	27.88%	41.43%	70.75%	51.37%	59.52%	73.24%	41.60%	53.06%
22	2	494620	28.36s	78.95%	28.85%	42.25%	71.57%	50.00%	58.87%	73.57%	41.20%	52.82%
21	1	382410	29.64s	73.17%	28.85%	41.38%	70.09%	51.37%	59.29%	70.95%	42.00%	52.76%
1	3	446824	26.45s	68.29%	26.92%	38.62%	71.70%	52.05%	60.32%	70.75%	41.60%	52.39%
47	3	707630	28.41s	73.17%	28.85%	41.38%	72.73%	49.32%	58.78%	72.86%	40.80%	52.31%
13	1	295260	29.39s	74.36%	27.88%	40.56%	68.81%	51.37%	58.82%	70.27%	41.60%	52.26%
52	1	466802	28.53s	77.50%	29.81%	43.06%	70.30%	48.63%	57.49%	72.34%	40.80%	52.17%
45	3	381976	28.52s	72.50%	27.88%	40.28%	68.81%	51.37%	58.82%	69.80%	41.60%	52.13%
31	1	417522	32.80s	66.67%	28.85%	40.27%	69.44%	51.37%	59.06%	68.63%	42.00%	52.11%
34	1	518364	29.48s	71.79%	26.92%	39.16%	71.84%	50.68%	59.44%	71.83%	40.80%	52.04%
4	3	366366	28.92s	72.50%	27.88%	40.28%	69.81%	50.68%	58.73%	70.55%	41.20%	52.02%
7	2	295036	32.22s	63.04%	27.88%	38.67%	70.37%	52.05%	59.84%	68.18%	42.00%	51.98%
26	3	559398	27.77s	78.38%	27.88%	41.13%	70.59%	49.32%	58.06%	72.66%	40.40%	51.93%
29	0	491288	29.42s	82.86%	27.88%	41.73%	69.23%	49.32%	57.60%	72.66%	40.40%	51.93%
30	1	576450	32.48s	70.73%	27.88%	40.00%	70.87%	50.00%	58.63%	70.83%	40.80%	51.78%
35	2	398342	35.08s	70.00%	26.92%	38.89%	69.44%	51.37%	59.06%	69.59%	41.20%	51.76%
10	1	418768	34.48s	67.44%	27.88%	39.46%	69.16%	50.68%	58.50%	68.67%	41.20%	51.50%
48	2	484680	32.83s	65.91%	27.88%	39.19%	69.81%	50.68%	58.73%	68.67%	41.20%	51.50%
40	1	550592	32.14s	76.32%	27.88%	40.85%	68.57%	49.32%	57.37%	70.63%	40.40%	51.40%
3	2	388262	29.58s	76.32%	27.88%	40.85%	66.97%	50.00%	57.25%	69.39%	40.80%	51.39%
42	2	540232	35.13s	64.29%	25.96%	36.99%	71.43%	51.37%	59.76%	69.39%	40.80%	51.39%
14	2	353290	29.50s	69.05%	27.88%	39.73%	66.67%	50.68%	57.59%	67.32%	41.20%	51.12%
23	1	524846	31.22s	72.50%	27.88%	40.28%	69.61%	48.63%	57.26%	70.42%	40.00%	51.02%
24	4	314244	27.53s	65.91%	27.88%	39.19%	68.87%	50.00%	57.94%	68.00%	40.80%	51.00%
6	2	437948	28.84s	80.00%	26.92%	40.29%	69.31%	47.95%	56.68%	72.06%	39.20%	50.78%

36	2	710164	34.11s	80.00%	26.92%	40.29%	69.31%	47.95%	56.68%	72.06%	39.20%	50.78%
33	3	478282	29.03s	68.29%	26.92%	38.62%	68.22%	50.00%	57.71%	68.24%	40.40%	50.75%
43	2	488544	28.13s	78.38%	27.88%	41.13%	69.00%	47.26%	56.10%	71.53%	39.20%	50.65%
46	2	506842	31.70s	82.86%	27.88%	41.73%	67.65%	47.26%	55.65%	71.53%	39.20%	50.65%
39	2	449302	32.58s	73.17%	28.85%	41.38%	68.32%	47.26%	55.87%	69.72%	39.60%	50.51%
27	0	426496	27.56s	76.92%	28.85%	41.96%	69.07%	45.89%	55.14%	71.32%	38.80%	50.26%
51	0	587790	29.84s	76.32%	27.88%	40.85%	68.69%	46.58%	55.51%	70.80%	38.80%	50.13%
37	1	529592	27.28s	70.00%	26.92%	38.89%	68.63%	47.95%	56.45%	69.01%	39.20%	50.00%
15	2	364672	26.55s	61.36%	25.96%	36.49%	68.22%	50.00%	57.71%	66.23%	40.00%	49.88%
20	0	596680	29.38s	75.00%	25.96%	38.57%	71.58%	46.58%	56.43%	72.52%	38.00%	49.87%
28	3	410032	24.27s	69.23%	25.96%	37.76%	70.00%	47.95%	56.91%	69.78%	38.80%	49.87%
49	0	625170	29.36s	80.00%	26.92%	40.29%	71.74%	45.21%	55.46%	74.02%	37.60%	49.87%
12	3	371154	27.23s	70.00%	26.92%	38.89%	67.31%	47.95%	56.00%	68.06%	39.20%	49.75%
44	1	583072	29.08s	77.78%	26.92%	40.00%	68.37%	45.89%	54.92%	70.90%	38.00%	49.48%
5	2	472108	28.13s	77.78%	26.92%	40.00%	70.21%	45.21%	55.00%	72.31%	37.60%	49.47%
50	2	624778	25.72s	75.68%	26.92%	39.72%	69.57%	43.84%	53.78%	71.32%	36.80%	48.55%
53	3	485352	26.64s	75.68%	26.92%	39.72%	64.95%	43.15%	51.85%	67.91%	36.40%	47.40%
0	11	38514	25.81s	53.85%	26.92%	35.90%	62.77%	40.41%	49.17%	59.59%	34.80%	43.94%

表 5.10 给出了模板 16 的特征组合方式。动词模型的最优试验结果在表 5.11 中给出。从表 5.11 中可见，名词模型中核心框架元素的 F 值明显低于非核心框架元素的 F 值，这再次说明了条件随机模型的训练偏执问题。

表 5.10 模板 16 对应的特征组合方式

编号	特征	选取方式	编号	特征	选取方式
1	词	$[-1, 1]$	14	关系标记	$[0, 0]$
2	词二元组	-	15	关系标记二元组	-
3	词性	$[-2, 2]$	16	中心词	$[0, 0]$
4	词性二元组	$[-2, 2]$	17	中心词二元组	$[-1, 1]$
5	位置	$[-1, 1]$	18	中心词性	$[0, 0]$
6	位置二元组	$[-2, 2]$	19	中心词性二元组	$[-2, 2]$
7	位置三元组	-	20	句法与中心词搭配	$[-1, 1]$
8	词与词性的搭配	$[-1, 1]$	21	关系与中心词搭配	$[0, 0]$
9	词与位置的搭配	$[0, 0]$	22	浅层句法路径	$[0, 0]$
10	词性与位置的搭配	$[0, 0]$	23	命名实体	$[0, 0]$
11	句法标记	$[-1, 1]$	24	浅层子范畴	-
12	句法标记二元组	$[-2, 2]$	25	是否简单句	-
13	句法标记三元组	$[-1, 1]$	26	目标词	$[0, 0]$

表 5.11 “陈述”框架名词模型的最优实验结果

名字	原来	预测	正确	P	R	F
medium	24	3	1	33.33%	4.17%	7.41%
msg	4	0	0	0.00%	0.00%	0.00%
spkr	76	34	28	82.35%	36.84%	50.91%
核心元素	104	37	29	78.38%	27.88%	41.13%
depic	0	0	0	0.00%	0.00%	0.00%
degr	6	0	0	0.00%	0.00%	0.00%
manr	57	53	39	73.58%	68.42%	70.91%
add	0	0	0	0.00%	0.00%	0.00%
top	70	48	36	75.00%	51.43%	61.02%
time	13	5	3	60.00%	23.08%	33.33%
cau	0	0	0	0.00%	0.00%	0.00%
非核心元素	146	106	78	73.58%	53.42%	61.90%
合计	250	143	107	74.83%	42.80%	54.45%

同样,本文对名词模型融合基本块信息前后,最优模型的各语义角色的F值(表5.11和表4.10)进行了对比,如图5.2所示。从图5.2中可以看出,加入基本块信息以后,各个语义角色的F值均有所下降。

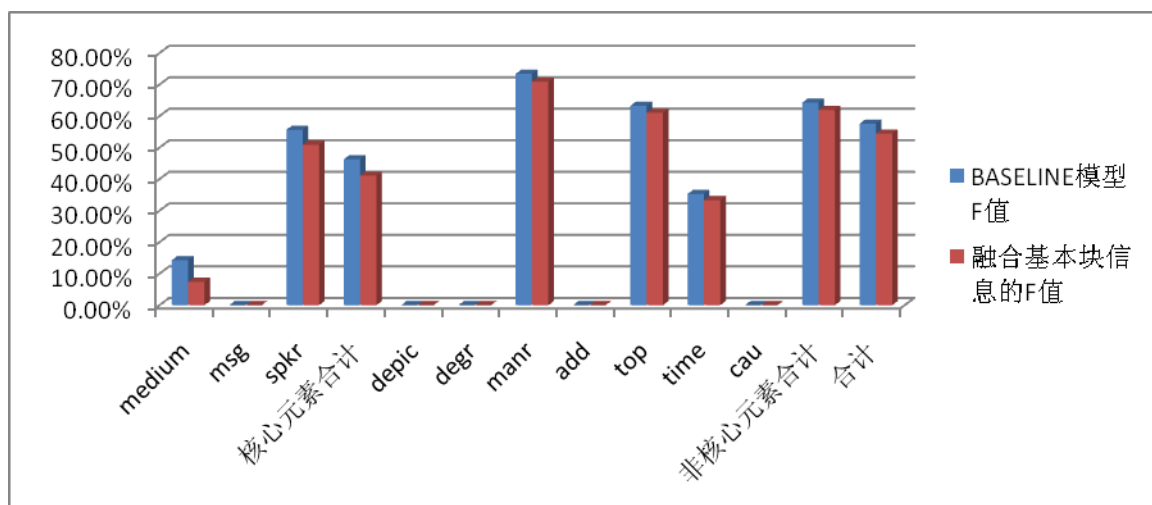


图 5.2 加入基本块信息前后名词模型各语义角色的F值的对比

5.4.3 实验分析

本文针对“陈述”框架下的动词词元模型和名词词元模型进行了特征的极差分析,分别见表5.12和表5.13。从中可以看出,动词模型中,“词/位置”特征的极差最大,为1.47%,而“句法标记三元组”特征的极差最小,仅为0.49%。名词模型中,“位置二元组”特征的极差最大,为1.41%,而“关系标记”特征的极差最小,为0.25%。根据极差分析的结果,每种特征的均值最大的水平可以被用来进行“算一算”

的实验, 本文不再给出相应的实验结果。

表 5.12 基于词的动词模型的极差分析结果

编号	特征类型	均值最大			均值最小			极差
		水平	窗口	均值	水平	窗口	均值	
1	词/位置	2	[-1, 1]	69.55%	0	-	68.08%	1.47%
2	位置	2	[-2, 2]	69.61%	0	[0, 0]	68.22%	1.40%
3	浅层路径	1	[0, 0]	69.38%	0	-	68.03%	1.35%
4	关系标记/中心词	2	[-1, 1]	69.30%	0	-	68.16%	1.14%
5	子范畴	1	[0, 0]	69.28%	0	-	68.14%	1.14%
6	句法标记/中心词	2	[-1, 1]	69.34%	0	-	68.21%	1.13%
7	位置三元组	2	[-2, 2]	69.42%	0	-	68.30%	1.12%
8	关系标记二元组	1	[-1, 1]	69.28%	0	-	68.17%	1.11%
9	词	1	[-1, 1]	69.26%	0	[0, 0]	68.17%	1.09%
10	位置二元组	2	[-2, 2]	69.43%	0	-	68.35%	1.08%
11	是否简单句	2	[0, 0]	69.26%	0	-	68.19%	1.08%
12	词性二元组	2	[-2, 2]	69.30%	0	-	68.24%	1.06%
13	词/词性	1	[0, 0]	69.23%	0	-	68.20%	1.03%
14	词性/位置	2	[-1, 1]	69.31%	0	-	68.41%	0.90%
15	中心词	2	[-2, 2]	69.17%	0	[0, 0]	68.33%	0.84%
16	命名实体	1	[0, 0]	69.20%	0	-	68.37%	0.84%
17	词性	1	[-1, 1]	69.17%	0	[0, 0]	68.34%	0.83%
18	关系标记	2	[-2, 2]	69.20%	0	[0, 0]	68.38%	0.82%
19	句法标记	1	[-1, 1]	69.22%	0	[0, 0]	68.42%	0.79%
20	句法标记二元组	1	[-1, 1]	69.17%	0	-	68.38%	0.79%
21	中心词二元组	1	[-1, 1]	69.25%	0	-	68.48%	0.77%
22	中心词性二元组	1	[-1, 1]	69.17%	0	-	68.41%	0.76%
23	词的二元组	1	[-1, 1]	69.12%	0	-	68.42%	0.71%
24	中心词性	2	[-2, 2]	69.14%	0	[0, 0]	68.43%	0.71%
25	句法标记三元组	1	[-1, 1]	69.10%	0	-	68.61%	0.49%

表 5.13 基于词的名词模型的极差分析结果

编号	特征类型	均值最大			均值最小			极差
		水平	窗口	均值	水平	窗口	均值	
1	位置二元组	2	[-2, 2]	52.08%	0	-	50.67%	1.41%
2	位置	2	[-2, 2]	52.02%	0	[0, 0]	50.64%	1.38%
3	句法标记/中心词	2	[-1, 1]	51.95%	0	-	50.64%	1.31%
4	词/词性	1	[0, 0]	52.04%	0	-	50.85%	1.19%
5	关系标记二元组	2	[-2, 2]	51.92%	0	-	50.81%	1.11%
6	词的二元组	1	[-1, 1]	51.99%	2	[-2, 2]	50.94%	1.05%
7	句法标记三元组	1	[-1, 1]	52.04%	0	-	51.02%	1.02%
8	词	1	[-1, 1]	51.85%	0	[0, 0]	50.89%	0.96%

9	中心词性	0	[0, 0]	51.81%	2	[-2, 2]	50.86%	0.94%
10	命名实体	1	[0, 0]	51.85%	0	-	51.00%	0.85%
11	词/位置	1	[0, 0]	51.77%	0	-	50.93%	0.84%
12	关系标记/中心词	2	[-1, 1]	51.78%	0	-	50.98%	0.80%
13	浅层路径	2	[0, 0]	51.96%	1	[0, 0]	51.21%	0.76%
14	句法标记	1	[-1, 1]	51.88%	2	[-2, 2]	51.16%	0.72%
15	中心词	1	[-1, 1]	51.69%	0	[0, 0]	51.07%	0.62%
16	子范畴	0	-	51.83%	2	[0, 0]	51.25%	0.59%
17	位置三元组	1	[-1, 1]	51.69%	0	-	51.13%	0.57%
18	词性/位置	2	[-1, 1]	51.71%	0	-	51.16%	0.55%
19	句法标记二元组	1	[-1, 1]	51.64%	2	[-2, 2]	51.17%	0.47%
20	词性	1	[-1, 1]	51.68%	2	[-2, 2]	51.23%	0.46%
21	词性二元组	2	[-2, 2]	51.64%	0	-	51.19%	0.45%
22	中心词二元组	0	-	51.73%	2	[-2, 2]	51.29%	0.43%
23	中心词性二元组	2	[-2, 2]	51.62%	0	-	51.20%	0.42%
24	是否简单句	1	[0, 0]	51.61%	0	-	51.30%	0.30%
25	关系标记	2	[-2, 2]	51.56%	0	[0, 0]	51.31%	0.25%

当然，词特征的稀疏程度和边界识别的精度仍然是分析模型的性能的重要因素，由于加入基本块信息后，模型的性能并没有显著地增长，因此，本文在这里不再给出相应的量化指标。除了词特征具有稀疏性外，模型中中心词特征和浅层路径特征也同样具有稀疏性，对于它们的分析仍旧可以使用对词特征的分析方法进行。

5.4.4 其它框架的实验结果

本文对其余的 24 个框架也进行了相应的实验，并将 25 个框架的实验结果汇总到一起，具体见表 5.14 所示。

表 5.14 25 个框架的基于词的标注模型的实验结果

框架名	模板	核心语义角色			非核心语义角色			所有语义角色		
		P	R	F	P	R	F	P	R	F
联想	14	49.38%	27.21%	35.09%	0.00%	0.00%	0.00%	49.38%	27.21%	35.09%
范畴化	2	78.57%	33.33%	46.81%	66.67%	23.08%	34.29%	76.92%	31.65%	44.84%
查看	9	70.89%	35.67%	47.46%	73.24%	34.67%	47.06%	72.00%	35.18%	47.26%
因果	10	79.10%	44.54%	56.99%	100.00%	6.67%	12.50%	79.41%	40.30%	53.47%
陈述_n	16	78.38%	27.88%	41.13%	73.58%	53.42%	61.90%	74.83%	42.80%	54.45%
发明	48	71.33%	53.97%	61.45%	100.00%	13.33%	23.53%	72.85%	44.18%	55.00%
非自主感知	11	70.20%	55.66%	62.09%	70.00%	22.34%	33.87%	70.18%	47.89%	56.93%
宗教信仰	48	78.79%	46.43%	58.43%	0.00%	0.00%	0.00%	78.79%	44.83%	57.14%
感受	19	54.17%	44.61%	48.92%	76.36%	69.42%	72.73%	62.95%	53.85%	58.04%
思想	18	78.38%	46.77%	58.59%	0.00%	0.00%	0.00%	78.38%	46.77%	58.59%
自主感知	48	75.36%	61.03%	67.44%	61.90%	19.26%	29.38%	73.90%	50.98%	60.34%
思考	48	71.62%	56.68%	63.28%	91.67%	27.50%	42.31%	73.12%	51.54%	60.47%

证明	7	63.51%	61.84%	62.67%	100.00%	37.50%	54.55%	66.25%	57.61%	61.63%
适宜性	8	69.23%	57.69%	62.94%	0.00%	0.00%	0.00%	69.23%	55.56%	61.64%
相信	8	84.38%	61.36%	71.05%	88.89%	24.62%	38.55%	85.09%	49.24%	62.38%
外观	21	80.00%	54.32%	64.71%	66.67%	28.57%	40.00%	79.31%	52.27%	63.01%
记忆	41	79.38%	59.69%	68.14%	93.75%	35.29%	51.28%	81.42%	53.64%	64.67%
获知	9	78.57%	62.03%	69.33%	84.00%	37.50%	51.85%	79.15%	57.76%	66.79%
鲜明性	19	79.87%	66.48%	72.56%	100.00%	3.45%	6.67%	80.00%	57.69%	67.04%
计划	10	75.00%	64.29%	69.23%	75.00%	54.55%	63.16%	75.00%	63.16%	68.57%
拥有	41	80.65%	60.61%	69.20%	0.00%	0.00%	0.00%	80.65%	59.88%	68.73%
代表	7	81.54%	63.10%	71.14%	100.00%	37.50%	54.55%	82.35%	60.87%	70.00%
陈述 _v	38	78.83%	68.97%	73.57%	80.93%	46.25%	58.86%	79.18%	63.63%	70.56%
知觉特征	15	86.89%	62.35%	72.60%	88.89%	67.47%	76.71%	87.57%	64.03%	73.97%
属于某类	10	88.33%	71.62%	79.10%	0.00%	0.00%	0.00%	88.33%	69.74%	77.94%
包含	4	90.55%	88.46%	89.49%	0.00%	0.00%	0.00%	90.55%	88.46%	89.49%
总计		76.24%	58.91%	66.46%	78.95%	39.45%	52.61%	76.68%	54.41%	63.65%

图 5.3 中给出了 25 个框架的基线模型和基于词的模型的总体 F 值的对比。从中可以看出，使用词语作为标注单位，基本块信息的加入使得模型的总体 F 值有 1% 的提高，并且很多框架加入基本块信息后，总体 F 值也有了稍微提高。

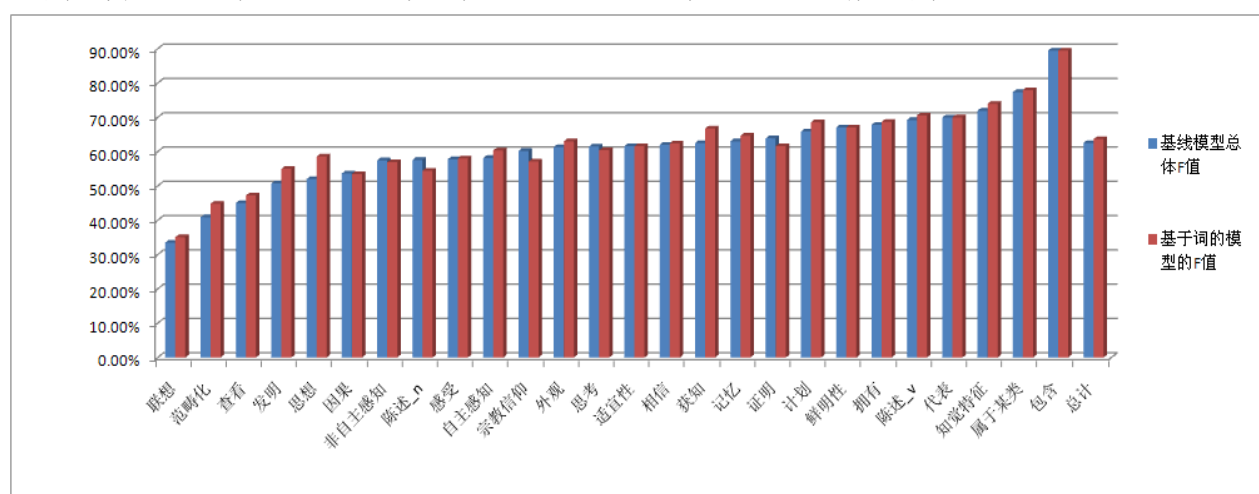


图 5.3 25 个框架的基线模型和基于词的模型的总体 F 值的对比

5.5 基于基本块的模型

本节使用基本块作为标注单位对汉语框架语义角色进行了标注。下面，本文首先给出对模型训练时的一些处理策略。在此基础上，阐述了模型所使用的特征信息和最优模板的选取方法。最后，本文给出了模型的详细实验结果以及相应的实验分析。

5.5.1 模型预处理

模型训练时,由于一些语义角色块无法和自动标注的基本块边界对齐,可能会导致语义角色块的训练样本变少,加剧模型的标注偏执问题。考虑到这一点,本文在训练时将无法和语义角色块对齐的基本块按照相应的规则打散成相应的单词块和块外词。然后,使用条件随机场模型进行训练。预测时,由于事先无法得知语义角色块的边界,因此,本文不对测试集的基本块进行预处理。

5.5.2 模型的特征描述及选取

针对基于基本块的语义角色标注模型,本文分别使用了如下几种基本块层面的特征:

块中心词特征:指当前标注的基本块的中心词词条。

块中心词的词性特征:指当前标注的基本块的中心词的词性。

块句法标记特征:指当前标注的基本块的句法标记。

块关系标记特征:指当前标注的基本块的关系标记。

位置特征:当前基本块的中心词与句子的目标词之间的相对位置,即{L, R, T}。

目标词特征:当前基本块所在句子的目标词元。

线性句法路径特征:以当前标注的基本块的句法标记和谓词的句法标记作为终结符,由途径的句法标记进行连接的链式特征。

命名实体的特征:标注出当前块的中心词是否是命名实体。具体的命名实体需要通过词性的映射来决定。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、nsy=外国地名、ni=机构专名},TIME包括{nt=时间名词}。

浅层子范畴特征:描述谓词词元所在基本块周围的基本块组成的序列,窗口为[-1,1]。

是否为简单句(二元特征):使用"是"和"否"来表示当前句子是否为简单句。

表 5.15 中给出了所有的候选特征和每种特征的候选窗口。其中,句法标记特征、关系标记特征、中心词特征、中心词性特征、位置特征和目标词特征作为必选特征被加入到模型中。在此基础上,本文考虑了它们的二元组合特征以及三元组合特征。

表 5.15 基于基本块的语义角色标注模型的特征选取方式

编号	特征	窗口			
1	句法标记	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
2	句法标记二元组	-	[-1, 1]	[-2, 2]	[-3, 3]
3	关系标记	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
4	关系标记二元组	-	[-1, 1]	[-2, 2]	[-3, 3]
5	中心词	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
6	中心词二元组	-	[-1, 1]	[-2, 2]	[-3, 3]
7	中心词性	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
8	中心词性二元组	-	[-1, 1]	[-2, 2]	[-3, 3]
9	位置	[0, 0]	[-1, 1]	[-2, 2]	[-3, 3]
10	位置二元组	-	[-1, 1]	[-2, 2]	[-3, 3]
11	浅层句法路径	-	[0, 0]		
12	命名实体	-	[0, 0]		
13	浅层子范畴	-	[0, 0]		
14	是否简单句	-	[0, 0]		
15	中心词三元组	-	[-1, 1]		
16	中心词性三元组	-	[-1, 1]		
17	句法标记三元组	-	[-1, 1]		
18	成分标记三元组	-	[-1, 1]		
19	位置三元组	-	[-1, 1]		
20	目标词	[0, 0]			

本文使用 PB 表 $PB_{48}(4^{12} \times 2^{11})$ 从候选的特征模板空间中挑选最好的实验模板。本文仅使用了该正交表中的 9 个 2 水平列，以及 10 个 4 水平列。

本文同样针对 25 个框架进行了相应的实验。由于篇幅有限，本文仅针对“陈述”框架给出详细的实验结果和分析。

5.5.3 模型的实验结果

下面，本文分别给出“陈述”框架的动词模型和名词模型的实验结果。

1. “陈述”框架动词模型的试验结果

表 5.16 给出了“陈述”框架的基于基本块的动词模型的 48 组实验结果。其中，模板 5 的实验结果最优，总体 F 值仅达到了 68.55%。这比 BASELINE 模型有着 1% 的下降，并且核心框架语义角色和非核心框架语义角色的总体 F 值也相应下降。

表 5.16 “陈述”框架的基于基本块的动词模型的 48 组实验结果

模板 编号	非法 序列	特征数	训练时间	核心框架语义角色			非核心框架语义角色			所有框架语义角色		
				P	R	F	P	R	F	P	R	F
5	1	1631520	465.20s	75.50%	66.96%	70.98%	80.42%	46.73%	59.11%	76.33%	62.21%	68.55%
4	5	2249860	431.17s	76.82%	66.82%	71.47%	80.18%	44.07%	56.88%	77.36%	61.47%	68.51%

9	3	1171640	518.25s	75.08%	66.82%	70.71%	78.40%	47.46%	59.13%	75.66%	62.27%	68.31%
35	5	1525400	473.20s	76.25%	66.89%	71.26%	77.31%	44.55%	56.53%	76.43%	61.64%	68.24%
27	0	2240360	505.66s	76.52%	66.67%	71.25%	76.37%	43.83%	55.69%	76.49%	61.30%	68.06%
20	4	996080	492.66s	74.79%	66.89%	70.62%	76.49%	46.49%	57.83%	75.09%	62.09%	67.98%
30	6	639560	507.19s	74.79%	67.34%	70.87%	70.04%	46.97%	56.23%	73.91%	62.55%	67.76%
7	4	436840	624.02s	74.92%	66.89%	70.68%	71.85%	46.97%	56.81%	74.35%	62.21%	67.74%
47	3	1203140	435.00s	75.60%	66.15%	70.56%	76.33%	45.28%	56.84%	75.72%	61.24%	67.72%
41	2	1295680	404.38s	75.77%	66.29%	70.71%	76.47%	44.07%	55.91%	75.88%	61.07%	67.68%
42	5	1226620	477.25s	76.02%	66.29%	70.83%	73.79%	44.31%	55.37%	75.63%	61.13%	67.61%
39	2	1667500	524.25s	75.27%	67.04%	70.92%	72.98%	43.83%	54.77%	74.88%	61.58%	67.58%
10	3	1936360	452.72s	75.79%	65.70%	70.39%	79.39%	43.83%	56.47%	76.38%	60.56%	67.56%
2	3	1262940	469.70s	74.85%	66.89%	70.65%	75.10%	43.83%	55.35%	74.90%	61.47%	67.52%
40	5	629720	497.44s	74.50%	66.29%	70.16%	76.00%	46.00%	57.32%	74.76%	61.53%	67.50%
29	3	982600	441.50s	75.13%	65.62%	70.06%	77.24%	46.00%	57.66%	75.49%	61.01%	67.49%
12	4	826160	428.77s	74.68%	65.62%	69.86%	79.17%	46.00%	58.19%	75.44%	61.01%	67.46%
24	4	1531500	467.09s	75.17%	66.00%	70.29%	76.76%	44.79%	56.57%	75.44%	61.01%	67.46%
16	5	1888160	455.86s	75.86%	65.70%	70.41%	77.83%	43.34%	55.68%	76.18%	60.44%	67.41%
3	4	1757700	390.66s	75.86%	65.70%	70.41%	78.67%	42.86%	55.49%	76.31%	60.33%	67.39%
17	4	1896260	458.06s	75.43%	65.77%	70.27%	80.09%	42.86%	55.84%	76.17%	60.39%	67.37%
13	2	1862700	432.44s	75.73%	65.25%	70.10%	81.00%	43.34%	56.47%	76.58%	60.10%	67.35%
11	4	2233940	472.89s	75.34%	66.59%	70.70%	76.79%	41.65%	54.00%	75.57%	60.73%	67.34%
32	2	1944820	463.33s	75.34%	65.92%	70.32%	76.62%	42.86%	54.97%	75.55%	60.50%	67.19%
19	5	763360	554.97s	74.94%	66.29%	70.35%	72.16%	44.55%	55.09%	74.45%	61.18%	67.17%
14	5	615900	542.92s	74.03%	66.59%	70.11%	73.52%	45.04%	55.86%	73.94%	61.53%	67.16%
34	5	772620	498.75s	74.44%	66.29%	70.13%	74.69%	44.31%	55.62%	74.48%	61.13%	67.15%
23	7	1960940	423.64s	75.36%	66.00%	70.37%	76.89%	41.89%	54.23%	75.61%	60.33%	67.11%
15	5	1152900	358.24s	74.89%	65.48%	69.87%	77.78%	44.07%	56.26%	75.37%	60.44%	67.09%
46	4	1099220	509.30s	74.75%	65.62%	69.89%	74.80%	45.28%	56.41%	74.76%	60.84%	67.09%
21	2	1812000	401.55s	75.76%	65.10%	70.03%	79.45%	42.13%	55.06%	76.35%	59.70%	67.01%
22	1	2234820	449.88s	75.13%	65.62%	70.06%	77.53%	42.62%	55.00%	75.52%	60.22%	67.00%
33	5	1688280	522.05s	75.24%	65.33%	69.93%	77.97%	42.86%	55.31%	75.68%	60.05%	66.96%
25	6	1299460	461.41s	74.87%	65.62%	69.94%	75.97%	42.86%	54.80%	75.05%	60.27%	66.86%
45	2	2297980	453.06s	75.56%	64.88%	69.82%	80.66%	41.40%	54.72%	76.35%	59.36%	66.79%
31	3	1676020	471.56s	76.02%	64.88%	70.01%	75.98%	42.13%	54.21%	76.02%	59.53%	66.77%
8	2	1270600	480.47s	75.17%	64.88%	69.65%	77.73%	43.10%	55.45%	75.59%	59.76%	66.75%
44	2	1718880	495.41s	74.85%	65.10%	69.64%	78.38%	42.13%	54.80%	75.41%	59.70%	66.65%
1	4	2512140	404.38s	75.79%	64.06%	69.44%	80.48%	40.92%	54.25%	76.52%	58.62%	66.39%
28	5	1019440	391.75s	74.14%	65.48%	69.54%	73.44%	42.86%	54.13%	74.02%	60.16%	66.37%
18	5	1835980	379.30s	74.89%	64.14%	69.10%	79.72%	41.89%	54.92%	75.66%	58.91%	66.24%
26	6	2534800	357.34s	75.44%	64.21%	69.37%	78.24%	40.92%	53.74%	75.88%	58.74%	66.22%
36	5	2211460	508.69s	74.46%	64.43%	69.09%	76.07%	43.10%	55.02%	74.73%	59.42%	66.20%
43	7	1976400	393.92s	75.83%	64.21%	69.54%	75.34%	40.68%	52.83%	75.75%	58.68%	66.13%
38	6	1941300	390.72s	74.16%	64.06%	68.74%	80.18%	42.13%	55.24%	75.11%	58.91%	66.03%

6	3	1280060	404.56s	73.24%	64.14%	68.39%	78.90%	41.65%	54.52%	74.12%	58.85%	65.61%
37	3	597040	564.94s	73.46%	64.66%	68.78%	71.72%	42.37%	53.27%	73.16%	59.42%	65.58%
0	17	92540	361.44s	61.24%	46.21%	52.67%	58.06%	34.87%	43.57%	60.62%	43.54%	50.68%

表 5.17 给出了模板 5 的最优特征组合方式。模板 5 对应的各语义角色的详细的实验结果在表 5.18 中给出。为了直观地分析当前语义角色标注模型的性能，图 5.4 中给出了它与“陈述”框架动词 BASELINE 模型（表 4.7）实验结果的对比。

表 5.17 模板 5 的特征组合方式

编号	特征	选取方式	编号	特征	选取方式
1	句法标记	[0, 0]	11	浅层句法路径	[0, 0]
2	句法标记二元组	[-1, 1]	12	命名实体	-
3	关系标记	[-3, 3]	13	浅层子范畴	-
4	关系标记二元组	[-3, 3]	14	是否简单句	[0, 0]
5	中心词	[-1, 1]	15	中心词三元组	[-1, 1]
6	中心词二元组	[-2, 2]	16	中心词性三元组	[-1, 1]
7	中心词性	[-2, 2]	17	句法标记三元组	-
8	中心词性二元组	-	18	成分标记三元组	-
9	位置	[-3, 3]	19	位置三元组	-
10	位置二元组	-	20	目标词特征	[0, 0]

表 5.18 模板 5 对应的动词模型的详细实验结果

名字	原来	预测	正确	P	R	F
medium	280	201	170	84.58%	60.71%	70.69%
msg	647	631	461	73.06%	71.25%	72.14%
spkr	417	360	269	74.72%	64.51%	69.24%
核心元素合计	1344	1192	900	75.50%	66.96%	70.98%
depic	25	16	6	37.50%	24.00%	29.27%
degr	1	0	0	0.00%	0.00%	0.00%
manr	101	43	32	74.42%	31.68%	44.44%
add	138	106	98	92.45%	71.01%	80.33%
top	14	1	0	0.00%	0.00%	0.00%
time	133	74	57	77.03%	42.86%	55.07%
cau	1	0	0	0.00%	0.00%	0.00%
非核心元素合计	413	240	193	80.42%	46.73%	59.11%
合计	1757	1432	1093	76.33%	62.21%	68.55%

从图 5.4 中可以看出，使用基本块作为标注单位，模型的总体 F 值较 BASELINE 模型有少量的下降。从各个语义角色来看，medium、spkr 和 time 等语义角色的 F 值有些许下降，而其它语义角色的 F 值有些许上升，但是，改变幅度都不显著。

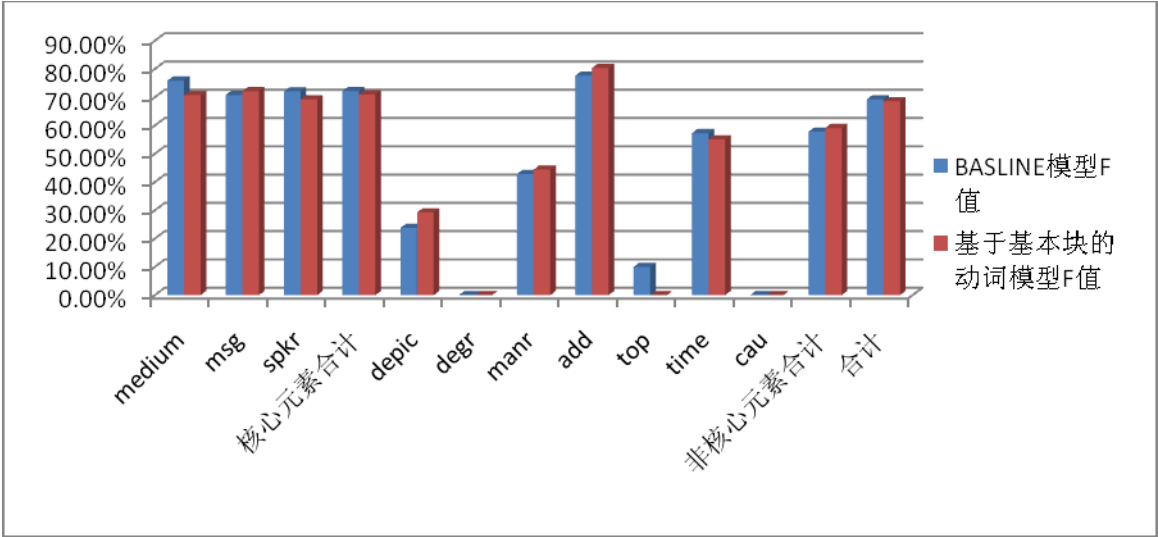


图 5.4 “陈述”框架动词词元的 BASELINE 模型和基于基本块模型实验结果的对比

2. “陈述”框架名词模型的实验结果

表 5.19 给出了“陈述”框架基于基本块的名词词元模型的 48 组实验结果。从中可以看出，模板 12 的实验结果最好，总体 F 值达到了 52.21%，较 BASELINE 模型有 5.37% 的下降。这种下降很大程度上是由于训练样本不充分和基本块的错误信息造成的。

表 5.19 “陈述”框架的基于基本块的名词模型的 48 组实验结果

模板 编号	非法 序列	特征数	训练时间	核心框架语义角色			非核心框架语义角色			所有框架语义角色		
				P	R	F	P	R	F	P	R	F
12	2	151410	23.03s	55.32%	25.00%	34.44%	68.38%	54.79%	60.84%	64.63%	42.40%	51.21%
7	5	104958	32.77s	56.86%	27.88%	37.42%	64.71%	52.74%	58.11%	62.35%	42.40%	50.48%
5	3	289310	27.11s	53.49%	22.12%	31.29%	71.30%	52.74%	60.63%	66.23%	40.00%	49.88%
46	1	218274	26.16s	47.83%	21.15%	29.33%	70.18%	54.79%	61.54%	63.75%	40.80%	49.76%
30	3	150080	28.61s	54.35%	24.04%	33.33%	66.38%	52.74%	58.78%	62.96%	40.80%	49.51%
4	2	400204	28.89s	51.11%	22.12%	30.87%	70.64%	52.74%	60.39%	64.94%	40.00%	49.50%
8	2	226646	25.78s	54.35%	24.04%	33.33%	68.81%	51.37%	58.82%	64.52%	40.00%	49.38%
21	0	342202	24.33s	60.98%	24.04%	34.48%	68.87%	50.00%	57.94%	66.67%	39.20%	49.37%
44	1	306306	29.69s	58.54%	23.08%	33.10%	70.19%	50.00%	58.40%	66.90%	38.80%	49.11%
14	3	135366	26.09s	48.00%	23.08%	31.17%	66.67%	53.42%	59.32%	61.08%	40.80%	48.92%
9	3	235144	26.78s	53.33%	23.08%	32.21%	66.67%	52.05%	58.46%	62.89%	40.00%	48.90%
20	4	200060	26.81s	47.06%	23.08%	30.97%	68.14%	52.74%	59.46%	61.59%	40.40%	48.79%
41	1	251384	25.27s	50.00%	20.19%	28.77%	69.37%	52.74%	59.92%	64.05%	39.20%	48.64%
35	2	266420	29.91s	43.75%	20.19%	27.63%	70.91%	53.42%	60.94%	62.66%	39.60%	48.53%
47	5	224490	24.03s	49.02%	24.04%	32.26%	69.16%	50.68%	58.50%	62.66%	39.60%	48.53%
22	2	391132	27.70s	55.81%	23.08%	32.65%	68.22%	50.00%	57.71%	64.67%	38.80%	48.50%
34	2	177198	24.48s	55.81%	23.08%	32.65%	64.10%	51.37%	57.03%	61.88%	39.60%	48.29%

13	1	322280	28.58s	57.89%	21.15%	30.99%	70.59%	49.32%	58.06%	67.14%	37.60%	48.21%
29	2	211260	23.59s	50.00%	22.12%	30.67%	67.57%	51.37%	58.37%	62.42%	39.20%	48.16%
42	3	235214	28.45s	55.00%	21.15%	30.56%	67.89%	50.68%	58.04%	64.43%	38.40%	48.12%
24	2	280560	28.72s	53.85%	20.19%	29.37%	69.16%	50.68%	58.50%	65.07%	38.00%	47.98%
2	2	233730	26.48s	56.10%	22.12%	31.72%	66.36%	50.00%	57.03%	63.58%	38.40%	47.88%
15	2	232568	22.14s	58.97%	22.12%	32.17%	66.67%	49.32%	56.69%	64.63%	38.00%	47.86%
45	2	384426	25.77s	52.50%	20.19%	29.17%	69.16%	50.68%	58.50%	64.63%	38.00%	47.86%
10	1	353206	24.14s	56.76%	20.19%	29.79%	70.59%	49.32%	58.06%	66.91%	37.20%	47.81%
39	3	280140	28.63s	52.27%	22.12%	31.08%	66.97%	50.00%	57.25%	62.75%	38.40%	47.64%
16	4	339920	31.56s	43.90%	17.31%	24.83%	69.03%	53.42%	60.23%	62.34%	38.40%	47.52%
25	2	249130	25.86s	52.17%	23.08%	32.00%	66.67%	49.32%	56.69%	62.34%	38.40%	47.52%
36	2	406266	30.02s	57.50%	22.12%	31.94%	66.67%	47.95%	55.78%	64.14%	37.20%	47.09%
37	4	129388	27.39s	49.06%	25.00%	33.12%	63.16%	49.32%	55.38%	58.68%	39.20%	47.00%
19	7	144410	31.70s	47.17%	24.04%	31.85%	60.48%	51.37%	55.56%	56.50%	40.00%	46.84%
1	0	435204	26.66s	58.33%	20.19%	30.00%	69.70%	47.26%	56.33%	66.67%	36.00%	46.75%
31	2	306992	24.58s	51.16%	21.15%	29.93%	68.63%	47.95%	56.45%	63.45%	36.80%	46.58%
33	2	280686	29.80s	46.94%	22.12%	30.07%	66.06%	49.32%	56.47%	60.13%	38.00%	46.57%
6	3	242116	22.44s	58.54%	23.08%	33.10%	64.76%	46.58%	54.18%	63.01%	36.80%	46.46%
32	1	326830	27.39s	52.50%	20.19%	29.17%	68.63%	47.95%	56.45%	64.08%	36.40%	46.43%
17	1	323344	28.06s	47.50%	18.27%	26.39%	69.90%	49.32%	57.83%	63.64%	36.40%	46.31%
43	2	344106	22.69s	65.71%	22.12%	33.09%	67.71%	44.52%	53.72%	67.18%	35.20%	46.19%
3	4	326144	24.83s	50.00%	24.04%	32.47%	66.02%	46.58%	54.62%	60.78%	37.20%	46.15%
18	2	323904	24.33s	57.14%	23.08%	32.88%	66.67%	45.21%	53.88%	63.83%	36.00%	46.04%
11	0	382802	28.92s	50.00%	20.19%	28.77%	66.99%	47.26%	55.42%	62.07%	36.00%	45.57%
38	2	352758	24.69s	58.54%	23.08%	33.10%	63.73%	44.52%	52.42%	62.24%	35.60%	45.29%
27	1	369894	29.42s	45.45%	19.23%	27.03%	66.67%	47.95%	55.78%	60.40%	36.00%	45.11%
23	1	347354	23.45s	51.16%	21.15%	29.93%	63.81%	45.89%	53.39%	60.14%	35.60%	44.72%
40	4	133630	24.81s	45.28%	23.08%	30.57%	60.91%	45.89%	52.34%	55.83%	36.40%	44.07%
28	3	197008	22.34s	42.31%	21.15%	28.21%	62.04%	45.89%	52.76%	55.62%	35.60%	43.41%
26	3	443940	22.98s	61.76%	20.19%	30.43%	62.89%	41.78%	50.21%	62.60%	32.80%	43.04%
0	13	18774	28.00s	48.78%	19.23%	27.59%	61.80%	37.67%	46.81%	57.69%	30.00%	39.47%

表 5.20 给出了模板 12 对应的特征组合方式。模板 12 对应的详细的实验结果在表 5.21 中给出。

表 5.20 模板 12 的特征组合方式

编号	特征	选取方式	编号	特征	选取方式
1	句法标记	$[-1, 1]$	11	浅层句法路径	-
2	句法标记二元组	$[-1, 1]$	12	命名实体	-
3	关系标记	$[-1, 1]$	13	浅层子范畴	-
4	关系标记二元组	$[-1, 1]$	14	是否简单句	-
5	中心词	$[-1, 1]$	15	中心词三元组	-
6	中心词二元组	$[-1, 1]$	16	中心词性三元组	-

7	中心词性	$[-1, 1]$	17	句法标记三元组	-
8	中心词性二元组	$[-1, 1]$	18	成分标记三元组	-
9	位置	$[-1, 1]$	19	位置三元组	-
10	位置二元组	$[-1, 1]$	20	目标词特征	$[0, 0]$

表 5.21 模板 12 对应的动词模型的详细实验结果

名字	原来	预测	正确	P	R	F
medium	24	5	3	60.00%	12.50%	20.69%
msg	4	0	0	0.00%	0.00%	0.00%
spkr	76	42	23	54.76%	30.26%	38.98%
核心元素合计	104	47	26	55.32%	25.00%	34.44%
depic	0	0	0	0.00%	0.00%	0.00%
degr	6	0	0	0.00%	0.00%	0.00%
manr	57	57	42	73.68%	73.68%	73.68%
add	0	0	0	0.00%	0.00%	0.00%
top	70	54	35	64.81%	50.00%	56.45%
time	13	6	3	50.00%	23.08%	31.58%
cau	0	0	0	0.00%	0.00%	0.00%
非核心元素合计	146	117	80	68.38%	54.79%	60.84%
合计	250	164	106	64.63%	42.40%	51.21%

图 5.5 给出了“陈述”框架名词词元的 BASELINE 模型和基于基本块模型的实验结果的对比。从中可以看出，使用基本块作为标注单位，绝大多数语义角色的 F 值较原有的模型有着大幅度下降。

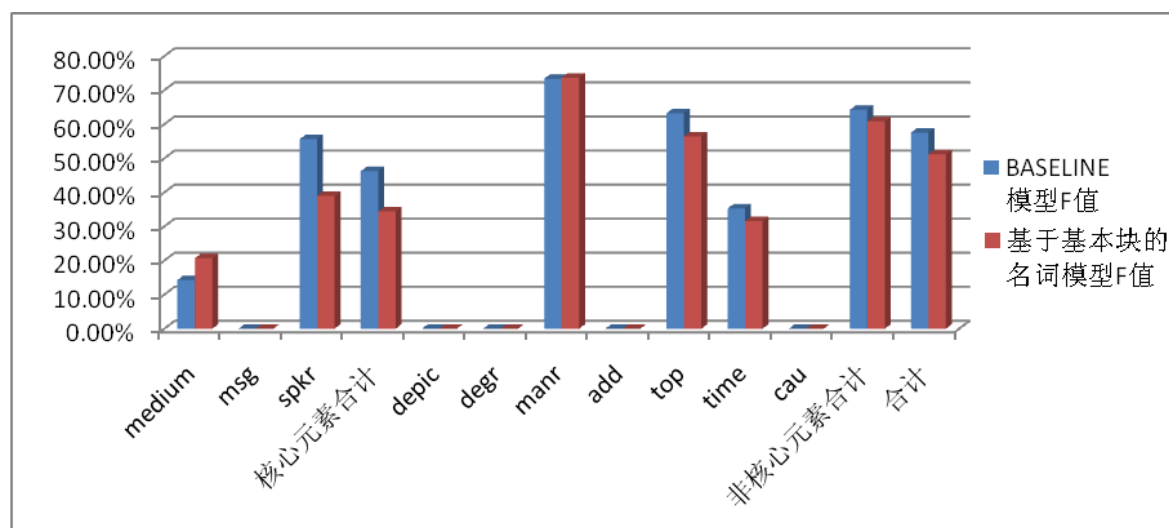


图 5.5 “陈述”框架名词词元的 BASELINE 模型和基于基本块模型实验结果的对比

5.5.4 实验分析

表 5.22 和表 5.23 中分别给出了基于基本块的动词模型和名词模型的极差分析

结果。从中可以看出，动词模型中，“位置”特征的重要程度最高，极差为 1.95%，而“成分标记三元组”特征的重要程度最低，极差仅为 0.34%。名词模型中，“位置”特征的重要程度最高，极差为 2.36%，而“成分标记三元组”特征的重要程度最低，极差为 0.01%。

表 5.22 基于基本块的动词模型的极差分析结果

编号	特征类型	均值最大			均值最小			极差
		水平	窗口	均值	水平	窗口	均值	
1	位置	3	[-3, 3]	67.49%	0	[0, 0]	65.54%	1.95%
2	中心词	2	[-2, 2]	67.36%	0	[0, 0]	65.51%	1.85%
3	位置二元组	3	[-3, 3]	67.42%	0	-	65.62%	1.81%
4	句法标记二元组	1	[-1, 1]	67.48%	0	-	65.77%	1.71%
5	关系标记二元组	2	[-2, 2]	67.26%	0	-	65.74%	1.52%
6	关系标记	1	[-1, 1]	67.29%	0	[0, 0]	65.85%	1.44%
7	中心词性二元组	1	[-1, 1]	67.27%	0	-	65.96%	1.31%
8	中心词性	1	[-1, 1]	67.18%	0	[0, 0]	65.90%	1.28%
9	浅层句法路径	1	[0, 0]	67.47%	0	-	66.19%	1.28%
10	中心词二元组	1	[-1, 1]	67.26%	0	-	66.01%	1.25%
11	句法标记	2	[-2, 2]	67.21%	0	[0, 0]	66.03%	1.18%
12	是否简单句	1	[0, 0]	67.22%	0	-	66.44%	0.78%
13	中心词三元组	1	[-1, 1]	67.20%	0	-	66.46%	0.74%
14	浅层子范畴	1	[0, 0]	67.18%	0	-	66.48%	0.70%
15	命名实体	1	[0, 0]	67.16%	0	-	66.50%	0.67%
16	句法标记三元组	1	[-1, 1]	67.14%	0	-	66.52%	0.62%
17	中心词性三元组	1	[-1, 1]	67.13%	0	-	66.52%	0.61%
18	位置三元组	1	[-1, 1]	67.14%	0	-	66.52%	0.61%
19	成分标记三元组	1	[-1, 1]	67.00%	0	-	66.66%	0.34%

表 5.23 基于基本块的名词模型的极差分析结果

编号	特征类型	均值最大			均值最小			极差
		水平	窗口	均值	水平	窗口	均值	
1	位置	3	[-3, 3]	48.29%	0		45.94%	2.36%
2	中心词	2	[-2, 2]	48.21%	0	[0, 0]	46.20%	2.01%
3	句法标记二元组	1	[-1, 1]	48.54%	0	-	46.88%	1.66%
4	中心词二元组	1	[-1, 1]	48.07%	3	[-3, 3]	46.53%	1.53%
5	位置二元组	3	[-3, 3]	47.93%	0		46.48%	1.44%
6	中心词性二元组	3	[-3, 3]	47.91%	0		46.48%	1.43%
7	关系标记二元组	1	[-1, 1]	48.08%	0	-	46.90%	1.17%
8	句法标记	1	[-1, 1]	47.86%	2	[-2, 2]	46.76%	1.10%
9	关系标记	2	[-2, 2]	47.73%	0	[0, 0]	46.63%	1.10%

10	浅层子范畴	0	-	47.87%	1	[0, 0]	46.83%	1.04%
11	命名实体	1	[0, 0]	47.85%	0	-	46.86%	0.99%
12	中心词三元组	1	[-1, 1]	47.63%	0	-	47.08%	0.55%
13	中心词性	1	[-1, 1]	47.50%	3	[-3, 3]	47.02%	0.48%
14	中心词性三元组	1	[-1, 1]	47.48%	0	-	47.23%	0.24%
15	是否简单句	1	[0, 0]	47.46%	0	-	47.24%	0.22%
16	位置三元组	1	[-1, 1]	47.45%	0	-	47.25%	0.20%
17	浅层句法路径	1	[0, 0]	47.44%	0	-	47.26%	0.18%
18	句法标记三元组	0	-	47.41%	1	[-1, 1]	47.30%	0.11%
19	成分标记三元组	0	-	47.36%	1	[-1, 1]	47.35%	0.01%

从“陈述”框架中基于基本块的动词和名词模型来看，使用基本块作为标注单位进行语义角色标注，模型的性能显著下降。本文认为产生这种现象的原因主要有两个：第一，一部分语义角色块的边界和基本块的边界无法对齐导致该语义角色块无法正确识别。第二，自动标注的基本块信息中含有大量的噪声信息，使得这种级联式模型的错误积累加剧。

5.5.5 其它框架的实验结果

本文使用基于基本块的标注模型对 25 个框架进行了自动语义角色标注，并将 25 个框架的实验结果汇总到一起，具体见表 5.24。

表 5.24 25 个框架的基于基本块的标注模型的实验结果

框架名	模板	核心语义角色			非核心语义角色			所有语义角色		
		P	R	F	P	R	F	P	R	F
联想	22	54.93%	26.53%	35.78%	0.00%	0.00%	0.00%	54.93%	26.53%	35.78%
范畴化	12	70.31%	34.09%	45.92%	63.64%	26.92%	37.84%	69.33%	32.91%	44.64%
查看	41	64.71%	35.03%	45.45%	65.06%	36.00%	46.35%	64.88%	35.50%	45.89%
陈述_n	12	55.32%	25.00%	34.44%	68.38%	54.79%	60.84%	64.63%	42.40%	51.21%
发明	2	68.79%	51.32%	58.79%	75.00%	10.00%	17.65%	69.13%	41.37%	51.76%
因果	13	77.61%	43.70%	55.91%	100.00%	6.67%	12.50%	77.94%	39.55%	52.48%
非自主感知	35	68.44%	54.05%	60.40%	62.07%	19.15%	29.27%	67.77%	45.91%	54.73%
相信	7	77.45%	59.85%	67.52%	71.43%	15.38%	25.32%	76.72%	45.18%	56.87%
感受	2	50.27%	45.59%	47.81%	74.14%	71.07%	72.57%	59.47%	55.08%	57.19%
自主感知	12	68.89%	58.22%	63.10%	58.82%	22.22%	32.26%	67.64%	49.55%	57.20%
宗教信仰	27	80.00%	50.00%	61.54%	0.00%	0.00%	0.00%	80.00%	48.28%	60.22%
适宜性	20	67.16%	57.69%	62.07%	0.00%	0.00%	0.00%	67.16%	55.56%	60.81%
记忆	5	79.68%	57.75%	66.97%	83.33%	23.53%	36.70%	80.09%	49.27%	61.01%
证明	45	67.14%	61.84%	64.38%	100.00%	25.00%	40.00%	68.92%	55.43%	61.45%
思想	12	78.05%	51.61%	62.14%	0.00%	0.00%	0.00%	78.05%	51.61%	62.14%
思考	35	73.86%	60.43%	66.47%	81.82%	22.50%	35.29%	74.39%	53.74%	62.40%
获知	14	75.69%	62.03%	68.18%	74.07%	35.71%	48.19%	75.51%	57.45%	65.26%
鲜明性	44	78.38%	64.80%	70.95%	100.00%	3.45%	6.67%	78.52%	56.25%	65.55%

拥有	12	75.97%	59.39%	66.67%	0.00%	0.00%	0.00%	75.97%	58.68%	66.22%
陈述_v	5	75.50%	66.96%	70.98%	80.42%	46.73%	59.11%	76.33%	62.21%	68.55%
外观	39	79.10%	65.43%	71.62%	100.00%	14.29%	25.00%	79.41%	61.36%	69.23%
代表	27	80.88%	65.48%	72.37%	100.00%	37.50%	54.55%	81.69%	63.04%	71.17%
计划	7	78.87%	66.67%	72.26%	80.00%	72.73%	76.19%	79.01%	67.37%	72.73%
知觉特征	9	84.92%	62.94%	72.30%	84.85%	67.47%	75.17%	84.90%	64.43%	73.26%
属于某类	12	88.14%	70.27%	78.20%	0.00%	0.00%	0.00%	88.14%	68.42%	77.04%
包含	7	92.80%	89.23%	90.98%	0.00%	0.00%	0.00%	92.80%	89.23%	90.98%
总计		73.58%	58.18%	64.98%	74.30%	38.69%	50.88%	73.70%	53.67%	62.11%

图 5.6 中给出了 25 个框架的基线模型和基于基本块模型的总体 F 值的比较。从中可以看出，使用基本块作为标注单位，除个别框架外，大部分框架的标注 F 值有所下降。总体 F 值有 0.4% 的下降。

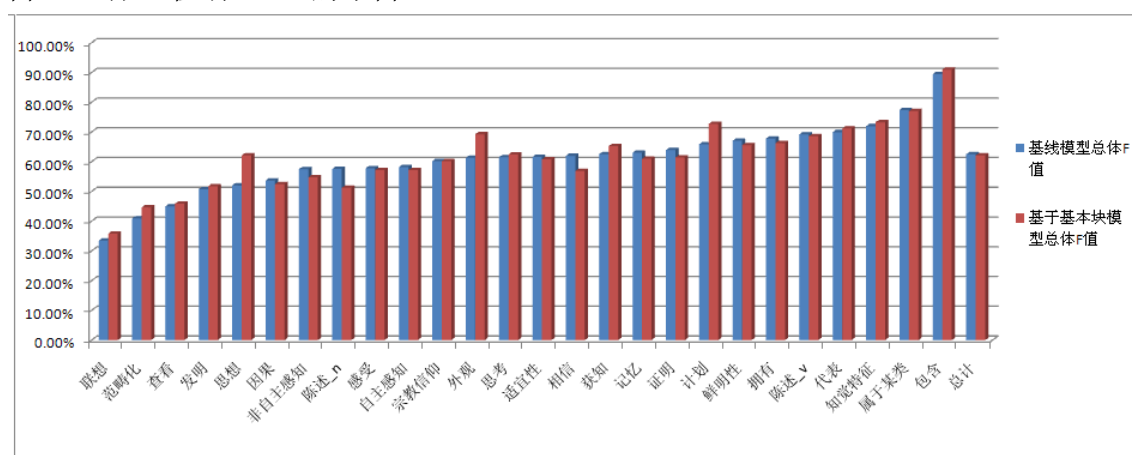


图 5.6 25 个框架的基线模型和基于基本块的模型的总体 F 值的对比

5.6 模型的比较与融合

有效地融合多种汉语框架语义角色标注模型的输出来提升语义角色的标注精度是目前语义角色标注的常用方法，也是未来统计自然语言处理的主流方法。但是，选取哪些模型来融合不同模型的语义角色自动标注结果，以及怎样融合这些模型的自动标注结果是多模型融合中必须考虑的问题。本节主要针对第一个问题提出了几种方法，并进行了初步的尝试。

针对两个汉语框架语义角色标注模型 A 和 B，如果 A 和 B 针对同一个测试集的自动标注结果有着很大的独立性或者负相关性，那么，融合 A 和 B 的自动预测结果可以显著地改善汉语框架语义角色标注模型的性能。对两个模型 A 和 B 自动预测结果的相关性的检验，可以使用如下两种方法进行检验。

第一种方法是 χ^2 一致性检验，这种检验方法常被使用到文本分类的特征提取技术^[79]中，也被借鉴到语义角色标注模型^[21]中来检验两个模型性能差异的显著程度。

这里, χ^2 一致性检验主要被用来检验两个模型的自动预测结果的相关性。两个语义角色标注模型的自动预测结果有四种情况, 如表 5.25 所示, 即正确的正例、正确的反例、第一类错误 (错误的正例)、以及第二类错误 (错误的反例)。这四种情况可以看做四个混淆集, 它们的数量分别为 N_{11} 、 N_{12} 、 N_{21} 和 N_{22} , 则 χ^2 的计算方法如(5-1)所示。

$$\chi^2 = \frac{(N_{11}N_{22} - N_{12}N_{21})^2 (N_{11} + N_{22} + N_{12} + N_{21})}{(N_{11} + N_{21})(N_{11} + N_{12})(N_{22} + N_{21})(N_{22} + N_{12})} \quad (5-1)$$

表 5.25 两个模型标注结果的分布情况

		B 系统		合计
		标注正确的个数	标注错的个数	
A 系统	标注正确的个数	N_{11}	N_{12}	$N_{1\bullet}$
	标注错的个数	N_{21}	N_{22}	$N_{2\bullet}$
合计		$N_{\bullet 1}$	$N_{\bullet 2}$	N

当 $p=0.05$ 时, $\chi^2 > \chi^2(1) = 3.481$, 两个系统的自动预测结果有 95% 的概率显著一致。

另一种方法是使用 Kappa 度量^[80]。Kappa 统计量是标注一致程度 (ITA, Inter tagger agreement) 的有效的度量, Gildea 和 Jurafsky 在文献[27]中使用该 Kappa 来度量 FrameNet 句子库的人工标注精度。它也可以被用来检验两个模型自动预测结果的一致性。对于表 5.25 中给出的两个模型的标注的分布, Kappa 度量可以通过观测一致 (Observed agreement) 和几率一致 (Chance agreement) 综合计算得到。Kappa 的具体计算方法如(5-2)所示。

$$Kappa = \frac{Observed\ agreement - Chance\ agreement}{1 - Chance\ agreement} \quad (5-2)$$

其中, (5-3)和(5-4)给出了观测一致 (Observed agreement) 和几率一致 (Chance agreement) 的计算方法。

$$Observed\ agreement = \frac{N_{11} + N_{22}}{N} \quad (5-3)$$

$$Chance\ agreement = \frac{N_{\bullet 1}N_{1\bullet} + N_{\bullet 2}N_{2\bullet}}{N} \quad (5-4)$$

Kappa 度量的取值范围为[-1,1]。当值为 1 时, 两个模型的实验结果完全一致。当

值为 0 时，两个模型的实验结果相互独立。当取值为-1 时，两个模型的实验结果完全负相关。

本文分别使用上述的卡方一致性检验和 Kappa 一致性度量对“陈述”框架的动词模型及名词模型的基线模型（记为 BL）、基于词的模型（记为 WB）和基于基本块的模型（记为 BB）进行了度量。本文仅考虑了三种模型对于语义角色块本身的标注情况，而不考虑块外词。本文以人工标注的测试集作为基准获得每两种自动标注模型的四个统计数据。对于“陈述”框架动词模型的检验结果见表 5.26。

表 5.26 “陈述”框架动词的三种模型预测结果一致性检验

模型 1	模型 2	N_{11}	N_{12}	N_{21}	N_{22}	χ^2	Kappa
BL	WB	1041	79	77	560	1147.197	0.808
BL	BB	985	135	108	529	870.481	0.703
WB	BB	1015	103	78	561	1067.963	0.779

从表 5.26 中可见， χ^2 值远远大于临界值，并且 N_{12} 和 N_{21} 的值较小，表现出强相关性。Kappa 的值也说明了这一点，并且 Kappa 三个值的顺序和 χ^2 是一致的。因此，“陈述”框架三个模型自动预测结果的融合意义不大。

“陈述”框架名词模型的检验结果见表 5.27。

表 5.27 “陈述”框架名词的三种模型预测结果一致性检验

模型 1	模型 2	N_{11}	N_{12}	N_{21}	N_{22}	χ^2	Kappa
BL	WB	98	16	9	127	159.488	0.797
BL	BB	97	17	9	127	156.359	0.789
WB	BB	96	11	10	133	171.507	0.828

表 5.27 中针对名词模型的结果显示出三种模型预测结果的正相关性很大。因此，对它们进行融合也是没有必要的。

本文对测试集中的其它 24 个框架也进行了相应的检验，相应的数据不再一一给出。从这些检验结果中可以发现，本文提出的三种标注模型的自动输出结果的正相关性很强，因此，对它们进行融合的意义不大。

5.7 本章小结

由于没有正确的基本块信息，本章将汉语基本块分析器自动产生的基本块信息融合到汉语框架语义角色标注模型中。本章认为汉语基本块应该是汉语语义角色块的

组成单位，汉语基本块和语义角色块边界的对应情况也粗略地说明了这一观点。针对这种关系，本章中提出了“基于词的模型”和“基于基本块的模型”两种汉语框架语义角色标注模型。

对于“基于词的模型”，本章将基本块形式化成词层面的若干特征加入到语义角色标注模型中，然后使用条件随机场模型对框架元素进行识别，并使用正交表对特征组合方式进行优化。实验结果表明，将基本块信息加入到基于词的模型中，模型的总体 F 值有了近 1% 的提高，总体 F 值达到 63.65%。

对于“基于基本块的模型”，本章直接将基本块组合形成汉语框架语义角色块。实验结果表明，使用基本块作为标注单位，总体 F 值较基线模型有 0.4% 的下降。本文认为，这种模型的性能更容易受到自动产生的基本块信息中含有的错误信息的影响。

本章的最后使用卡方一致性检验和 Kappa 度量两种方法来比较汉语框架语义角色的基线模型、基于词的模型和基于基本块的模型。实验结果表明，三种模型的预测结果有着较强的正相关性，因此，将这三种模型的自动预测结果进行融合的意义不是很大。

第6章 总结及展望

6.1 论文总结

本文主要研究了汉语框架语义角色的自动标注问题,其中,不包含通用语义角色。由于汉语完全句法分析器的性能不理想,本文将汉语框架语义角色自动标注看作是一种序列标注任务,并使用条件随机场模型对框架语义角色进行了自动的预测。

本文首先将模型建立在词语层面,并使用 IOB2 策略来实现语义角色的自动标注。考虑到同一个框架下不同词性的词元具有的语义角色搭配模式的分布可能不同,本文使用卡方检验来检验这种分布差异性是否显著,并且根据显著程度将一个框架的语义角色标注模型分裂成独立的几个模型。本文在所选取的 25 个框架上进行了实验。实验结果表明,仅使用词层面的信息,可以使 25 个框架的总体 F 值达到 63%左右,这是汉语框架语义角色标注的一个初步结果。

在此基础上,本文尝试将汉语基本块信息融合到框架语义角色标注模型中。鉴于汉语基本块是语义角色块的组成单位,本文提出了两种标注模型:“基于词的模型”和“基于基本块的模型”,并分别进行了实验。通过实验结果可以发现,基于词的模型中,基本块的加入对模型性能有所提升,但是并不显著。而在基于基本块的模型中,语义角色标注性能有所下降。本文认为,这种模型的性能更容易受到自动产生的基本块信息中含有的错误信息的影响。

为了进一步验证基本块信息的加入对模型自动预测结果的影响,本文分别使用卡方一致性检验和 Kappa 度量来比较汉语框架语义角色的基线模型、基于词的模型和基于基本块的模型。从实验结果可以看出,三种模型的预测结果的正相关性较强。这意味着基本块信息的加入并没有显著地改变基线模型的预测结果,并且将这三种模型的自动预测结果进行融合的意义不大。

另外,本文还进行了如下几方面的工作:

(1) 本文提出了 CFN 句法语义分析模型的概貌。这为本文今后工作的开展提供了一个清晰的蓝图。

(2) 本文中使用了正交表的方法对特征组合方式进行了选择和优化。实验结果表明,使用正交表进行特征的组合和优化可以有效地避免主观选取方法的盲目性,并且对提升语义角色的标注性能是十分有效的。

(3) 本文使用极差分析的方法给出了各类特征的重要性的一個度量。它可以

一个侧面反映出每类特征在模型中所起的作用。正交表的“算一算”方法可以建立在极差分析结果的基础上，进一步改善语义角色标注性能。但考虑到这种提升并不是必然的，因此本文没有给出相应的实验结果。

6.2 下一步研究计划

本文中对汉语框架语义角色标注的研究仍处于比较初步的阶段，今后需要进行的研究工作主要集中在以下几个方面：

目前来看，对汉语框架语义角色的短语类型和句法功能的自动标注研究是十分迫切的。它直接关系到汉语框架网络的句子的语义角色和句法之间的链接关系。并且，框架语义角色三层标记的自动识别为汉语句子的核心依存图的抽取提供了重要的基础。

汉语框架语义角色标注性能的进一步提高要求有更多语言学知识和语料库资源的引入。鉴于这一点，本文后续研究工作计划将汉语功能块信息以及汉语完全句法分析信息和汉语依存分析信息引入到语义角色标注模型中。值得一提的是汉语功能块信息，目前已经有相应的文献[44]表明汉语功能块的标注性能可以达到80%左右。并且，文献[72]和[81]的相关实验已经表明，汉语功能块信息的加入可以使得汉语框架语义角色及其句法功能的标注性能有着明显的提高。此外，为了避免词语等稀疏特征对模型性能的影响，考虑使用词语聚类、命名实体识别以及半监督的方法是十分有必要的。

通用语义角色标注是一个值得思考的问题。通用语义角色虽然是与词元有关的，但是，它所具有的语义信息在绝大多数（甚至所有）的框架中都是一致的。因此，是否可以将通用语义角色单独列为一个子任务进行研究，需要进一步进行理论和实验方面的论证。

目前的汉语框架语义角色模型仅仅使用了词层面和基本块层面的信息，这些信息虽然对找到词语与语义角色以及基本块与语义角色的对应关系产生了重要的作用。但是，对一个句子中出现的语义角色之间的搭配方式的约束较弱。实际上，将训练集或者汉语框架网络数据库中总结出来的语义角色搭配模型作为约束加入到模型中，可以对汉语句子的自动预测的语义角色搭配方式产生较强的约束。如何合理地引入这种信息是一件值得研究的课题。另外，借助框架网络中不同框架之间框架元素和框架元素之间的对应关系来扩充每个框架下的句子资源，从而缓解训练阶段的数据稀疏问题是值得思考的。

从机器学习的角度来研究汉语框架网路语义角色标注任务可以提供一种解决问题的新思路。从机器学习的侧面来看,需要进行的研究工作有如下几方面:

(1) 样本标注问题。目前汉语框架语义角色标注模型面临的瓶颈问题是训练样本比较少,那么,每一个框架下标注多少例句就足够了?这些例句应该随机地从CCL中自动的选取,还是人工有选择地进行选取?这些问题直接涉及到机器学习中样本个数求解和样本抽样问题,问题的求解必须借助于机器学习理论和自然语言处理的知识。

(2) 标注模型中框架元素个数对模型性能的影响。从本文的实验结果中可以看出,条件随机场模型对那些在训练集出现多的语义角色的标注性能较好,而对出现较少的语义角色标注性能比较差。这种现象被称之为“标注偏执”。另外,训练集中一个框架元素最少要出现多少次才值得训练?该问题的理论研究直接影响到汉语框架语义角色模型的调整和改进,例如,是否应该将通用语义角色单独进行识别?是否应该将语义角色边界识别任务和语义角色分类任务分成两个阶段进行?

(3) 本文所使用的序列标注的正交表特征选择方法目前仍处在初步的实验论证阶段,如何完善它的理论框架,并逐步形成一套基于正交表的特征分析方法,需要从统计学、机器学习和自然语言处理三个方面进行分析和探索。

(4) 半监督模型和无监督模型的引入是否会明显地改善汉语框架语义角色标注性能,需要进一步地进行实验方面和理论方面的论证。

(5) 条件随机场模型考虑到了序列中各个元素之间的相关性,并将整个序列作为训练样本。但是,在自然语言处理任务中,由于标注例句有限,使得条件随机场模型的训练样本个数严重不足,从而影响它的参数估计精度。相比之下,最大熵模型以序列中的每个元素作为训练样本,因此,它的训练样本集较为充足。所以,在序列标注的某些情况,最大熵比条件随机场模型的预测精度要高。另外,当最大熵模型的训练样本完全可分时,支撑向量机的方法可以达到理想的标注性能。因此,分别使用条件随机场模型、最大熵模型和支撑向量机模型进行汉语框架语义标注研究是十分必要的,而且比较这三种模型自动预测结果并制定相应的融合策略是一个十分有意义的课题。

目前,汉语框架语义角色标注的研究仍处在初步的阶段,但是随着对课题认识的深入和对相关问题的逐个解决,本文相信汉语框架语义角色标注器一定会研制成功。

参考文献

- [1] MARQUEZ L, CARRERAS X, C. LITKOWSKI K, STEVENSON S. Semantic Role Labeling: An Introduction to the Special Issue[J]. Computational Linguistics, 2008, **34**(2):145–159.
- [2] BAKER C F, FILLMORE C J, LOWE J B. The Berkeley FrameNet Project[C]. Proceedings of the ACL-Coling-1998, 1998, 86–90.
- [3] PALMER M, GILDEA D, KINGBURY P. The Proposition Bank: An Annotated Corpus of Semantic Roles.[J] Computational Linguistics, 2005, **31**(1):71–106.
- [4] MEYERS A, REEVES R, MACLEOD C. The Nombank Project: An Interim Report[C]. HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation, 2004, 24–31.
- [5] YOU L, LIU K. Building Chinese Framenet Database[C]. Conference on Natural Language Processing and Knowledge Engineering, 2005, 301–306.
- [6] XUE N, PALMER M. Annotating the Propositions in the Penn Chinese Treebank[C]. Proceedings of the Second SIGHAN Workshop on Chinese Language Processing, 2003, 47–54.
- [7] XUE N, XIA F, DONG CHIOU F. The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus[J]. Natural Language Engineering , 2005, **11**(2):207–238.
- [8] NARAYANAN S, HARABAGIU S. Question Answering based on Semantic Structures[C]. In Proceedings of COLING-2004, 2004, 22-29.
- [9] 李茹, 王文晶, 梁吉业. 基于汉语框架网的旅游信息问答系统设计[J]. 中文信息学报, 2009, **23**(2):34–40.
- [10] 宋小香, 李茹, 王文晶. 基于汉语框架网的中文问题分类[R]. 山西大学, 2009.
- [11] LAFFERTY J, MCCALLUM A, PEREIRA F. Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data[C]. In Proceedings of the 18th International Conf. on Machine Learning, 2001, 282-289.
- [12] CARRERAS X, MARQUEZ L. Introduction to the Conll-2004 Shared Task: Semantic Role Labeling[C]. HLT-NAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning, 2004, 89–97.
- [13] CARRERAS X, MARQUEZ L. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling[C]. Proceedings of CoNLL-2005, 2005, 152–164.

- [14] 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.
- [15] 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
- [16] ERK K, STRAPPARAVA C. SemEval-2010 Workshop[EB]. <http://semeval2.fbk.eu/semeval2.php>, 2010.
- [17] CHEN J, RAMBOW O. Use of Deep Linguistic Features for the Recognition and Labeling of Semantic Arguments[C]. Proceedings of EMNLP-2003, 2003.
- [18] THOMPSON C A, LEVY R, MANNING C D. A Generative Model for Semantic Role Labeling[C]. Proceedings of ECML-03, 2003, 22-26.
- [19] SWIER R S, STEVENSON S, Unsupervised Semantic Role Labelling[C]. Proceedings of EMNLP-2004, 2004.
- [20] HACIOGLU K. Semantic Role Labeling Using Dependency Trees[C]. Proceedings of COLING, 2004.
- [21] PRANDHAN S, HACIOGLU K, KRUGLER V, WARD W, MARTIN J, JURAFSKY D. Support Vector Learning for Semantic Argument Classification[J]. Machine Learning, 2005, **60**(1):11–39.
- [22] SURDEANU M, MARQUEZ L, CARRERAS X, COMAS P R. Combination Strategies for Semantic Role Labeling[J]. Journal of Artificial Intelligence Research, 2007, **29**:105-151.
- [23] PRADHAN S S, WARD W, MARTIN J H. Towards Robust Semantic Role[J]. Computational Linguistics, 2008, **34**(2):1-30.
- [24] LIU T, CHE W X, LI S. Semantic Role Labeling with Maximum Entropy Classifier[J]. Journal of Software, 2007, **18**(3):565–573.
- [25] 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.
- [26] CHE W X, LI Z H, HU Y X, LI Y Q, QIN B, LIU T, LI S. A Cascaded Syntactic and Semantic Dependency Parsing System[C]. Proceedings of the CoNLL-2008,

- 2008.
- [27] GILDEA D, JURAFSKY D. Automatic Labeling of Semantic Roles[J], Computational Linguistics, 2002, **28**(3):245-288.
- [28] XUE N, PALMER M. Automatic Semantic Role Labeling for Chinese Verbs[C]. Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, 2005.
- [29] XUE N. Labeling Chinese Predicates with Semantic Roles[J]. Computational Linguistics, 2008, **34**(2): 225-255.
- [30] YUAN Y L. the New Trend and Long Term Goal of Constructing Semantic Resource[J]. Journal of Chinese Information Processing, 2008, **22**(3):3-15.
- [31] 由丽萍, 刘开瑛. CFN句子标注规范[R]. 山西大学, 2008.
- [32] 刘焘. 汉语框架语义知识库信息组织与表示方法研究[D]. 山西大学, 2007.
- [33] JOHANSSON R, NUGUES P. LTH: Semantic Structure Extraction using Nonprojective Dependency Trees[C]. Proceedings of the 4th International Workshop on Semantic Evaluation, 2007: 227-230.
- [34] ERK K. Unknown Word Sense Detection as Outlier Detection. Proceedings of NAACL , 2006, 128-135.
- [35] MARKOU M, SINGH S, Novelty Detection: A Review Part1: Statistical Approaches[J]. Signal Processing, 2003, **83**(12):2481-2497.
- [36] MARKOU M, SINGH S. Novelty Detection: A Review Part 2: Neural Network based approaches[J], Signal Processing, 2003, **83**(12):2499-2521.
- [37] PROEFSCHRIFT. One-Class Classification[D], geboren te Ede, 2001.
- [38] ZHOU Z H, ZHANG M L. Multi-instance multi-label Learning with Application to Scene Classification[C]. Advances in Neural Information Processing Systems, 2007.
- [39] FILLMORE C J, RUPPENHOFER J, BAKER C F. FrameNet and Representing the Link between Semantic and Syntactic Relations[M]. Frontiers in Linguistics, 2002, **1**:19 - 59.
- [40] RABINER L R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition[C]. Proceedings of the IEEE, 1989:**77** (2):257-286.
- [41] MCCALLUM A, FREITAG D, PEREIRA F. Maximum Entropy Markov Models for Information Extraction and Segmentation[C]. Proceedings of ICML, 2000, 591-598.

- [42] DIETTERICH T G. Machine Learning for Sequential Data: A Review[C]. Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, 2002:15-30.
- [43] KOHAVI R, John G H. Wrappers for Feature Subset Selection[J]. Artificial Intelligence, 1997, 273-324.
- [44] 李国臣, 王瑞波, 李济洪. 基于条件随机场模型的汉语功能块自动标注[R]. 山西大学, 2008.
- [45] 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.
- [46] 李济洪, 王瑞波, 李国臣. 序列标注的正交表特征选择方法[R]. 山西大学, 2008
- [47] 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.
- [48] BERGER A L, VINCENT J, STEPHEN A, A Maximum Entropy Approach to Natural Language Processing[J]. Computational Linguistics, 1996, **22**(1):39-71.
- [49] LIU D C, NOCEDAL J. On the Limited Memory BFGS Method for Large-Scale Optimization[J]. Math Programming, 1989, **45**:503-528.
- [50] ALBERT A, ANDERSON J A. On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. Biometrika, 1984, **71**:1-10.
- [51] SPECKMAN P L, LEE J, SUN D. Existence of the MLE and Propriety of Posteriors for a General Multinomial Choice Model[J]. Duke University, 1999.
- [52] CHEN S F, RONALD R. A Gaussian Prior for Smoothing Maximum Entropy Models[R]. CMUCS, 1999.
- [53] 李济洪, 王瑞波, 李国臣. 汉语框架语义角色标注研究进展[R]. 山西大学, 2009.
- [54] CRAMER J S. Predictive Performance of the Binary Logit Model in Unbalanced Samples[J], The Statistician, 1999, **48**(1): 85-94
- [55] MCCALLUM A. Efficiently Inducing Features of Conditional Random Fields[C]. Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence, 2003.
- [56] ANDREW G, GAO J F. Scalable Training of L1-regularized Log-linear Models[C]. Proceedings of the 24th international conference on Machine learning, 2007, 33-40.
- [57] HASTIE T, TIBSHIRANI R, FRIEDMAN J. The Elements of Statistical Learning[M]. Springer Press, 2001.

- [58] MCCALLUM A. Efficiently inducing features of conditional random fields[C].
Proceeding of the Nineteenth Conference on Uncertainty in Artificial Intelligence, 2003.
- [59] KRUENGKRAI C, VANICH V S, ISAHARA H. A Conditional Random Field
Framework for Thai Morphological Analysis[C]. Proceedings of the Fifth International
Conference on Language Resources and Evaluation, 2006.
- [60] KUDO T, MATSUMOTO Y. Chunking with Support Vector Machines [C]. The Second
Annual Meeting of the North American Chapter of the Association for Computational
Linguistics [C], Pittsburgh, 2001: 192-199.
- [61] COLLINS M. Discriminative training methods for hidden Markov models: theory and
experiments with perceptron algorithms[C], Proceedings of the ACL-02 Conference on
Empirical methods in Natural Language Processing, 2002, 1-8.
- [62] PAL C, SUTTON C, MCCALLUM A. Sparse Forward-Backward using Minimum
Divergence Beams for Fast Training of Conditional Random Fields[C]. International
Conference on Acoustics, Speech, and Signal Processing, 2006.
- [63] SUTTON C, ROHANIMANESH K, McCallum A. Dynamic conditional random
fields: factorized probabilistic models for labeling and segmenting sequence data[C].
Proceedings of the twenty-first international conference on Machine learning, 2004.
- [64] FINKEL J R, GRENAGER T, MANNING C. Incorporating Non-local Information
into Information Extraction Systems by Gibbs Sampling[C], Proceedings of the 43rd
Annual Meeting on Association for Computational Linguistics, 2005, 363-370.
- [65] KRISHNAN V, MANNING C D. An Effective Two-stage Model for Exploiting
Non-local Dependencies in Named Entity Recognition[C]. Proceedings of the 21st
International Conference on Computational Linguistics and the 44th annual meeting of the
ACL, 2006, 1121-1128.
- [66] TANG J, HONG M, LI J, LIANG B. Tree-Structured Conditional Random Fields
for Semantic Annotation[C]. Proceedings of International Semantic Web Conference, 2006,
640-653.
- [67] ZHU J, NIE Z, WEN J, ZHANG B, MA W. Simultaneous Record Detection and
Attribute Labeling in Web Data Extraction[C]. Proceedings of KDD, 2006, 494-503.
- [68] KUDO. CRF++ toolkit[EB]. <http://crfpp.sourceforge.net/>, 2003.
- [69] OKAZAKI N. CRF Suite[EB]. <http://www.chokkan.org/software/crfsuite/>, 2007.

- [70] Phan X H. Flex CRF toolkit[EB].<http://flexcrfs.sourceforge.net/>, 2004.
- [71] MCCALLUM A, KACHITES A. MALLET [EB].<http://mallet.cs.umass.edu>, 2002.
- [72] 张国华. 汉语框架语义角色自动标注研究[D]. 山西大学, 2008.
- [73] PRADHAN S, HACIOGLU K, KRUGLER V, WARD W, MARTIN J, JURAFSKY D. Support vector learning for semantic argument classification[J]. *Machine Learning*, 2005, **6**(1): 11 – 39.
- [74] 刘开瑛. 汉语框架语义网构建现状[C]. 第四届全国学生计算语言学研讨会论文集, 2008.
- [75] 周强. 汉语基本块描述体系[J]. *中文信息学报*, 2007.
- [76] 周强. 基于规则的汉语基本块自动分析器[C]. 第七届中文信息处理国际会议论文集, 2007: 137–142.
- [77] 周强. 汉语树库构建使用手册[R]. 清华大学, 2005.
- [78] WU C F J, Hamada M. Experiments Planning, Analysis, and Parameter Design Optimization[M]. New York, 2003.
- [79] YANG Y. A Comparative Study on Feature Selection in Text Categorization [C]. *Proceeding of the Fourteenth International Conference on Machine Learning*, 1997, 412-420.
- [80] CARLETTA J. Assessing agreement on classification tasks: the kappa statistic[J]. *Computational Linguistics*, 1996, **22**(2): 249-254.
- [81] 尹晓丽. 基于条件随机场模型的通用语义角色自动标注研究[D]. 山西大学, 2008.

附录 A：山大 973 词性标注体系及其与清华大学词性映射关系

山大标记	含义	清华标记	含义	山大标记	含义	清华标记	含义
n	普通名词	n	名词	q	量词	q	量词
nt	时间名词	t	时间词	d	副词	d	副词
nd	方位名词	n	名词	r	代词	r	代词
nl	处所名词	s	处所词	p	介词	p	介词
nn	族名	n	名词	c	连词	c	连词
nh	人名	nP	指人专名	u	助词	u	助词
nhh	汉人名	nP	指人专名	e	叹词	e	叹词
nhf	汉姓氏	nP	指人专名	o	拟声词	o	象声词
nhg	汉名字	nP	指人专名	i	习用语	i	成语
nhy	音译名	nP	指人专名	in	名习用语	iN	名词性成语
nhr	日本人名	nP	指人专名	iv	动习用语	iV	谓词性成语
nhw	简称、笔名等	nP	指人专名	ia	形习用语	i	成语
ns	地名	nS	地点专名	ic	连习用语	i	成语
nsh	中国地名	nS	地点专名	j	简称	n	名词
nsy	外国地名	nS	地点专名	jn	名简称	n	名词
ni	机构专名	nO	机构名	jv	动简称	v	动词
nz	其它专名	nR	其他专名	ja	形简称	a	形容词
v	普通动词	v	动词	h	前接成分	h	前缀
vd	趋向动词	v	动词	k	后接成分	k	后缀
vl	系动词	vC	系动词	g	语素字	g	汉字
vu	能愿动词	v	动词	x	非语素字	x	任意字符串
a	形容词	a	形容词	w	其它	w	标点
aq	性质形容词	a	形容词	wp	标点符号	w	标点
as	状态形容词	a	形容词	ws	非汉字字符串	x	任意字符串
f	区别词	b	区别词	wu	其它未知符号	x	任意字符串
m	数词	m	数词				

附录 B: 实验用正交表及 PB 表

$L_{32}(4^9 \times 2^4)$													
条件号	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	4	3	2	1	4	4	4	1	1	2	2
2	2	1	1	1	3	3	3	2	4	1	1	1	1
3	3	1	3	4	1	3	1	4	2	1	2	2	1
4	4	1	2	2	4	1	2	2	2	1	2	1	2
5	1	2	3	2	3	3	2	3	1	1	1	2	2
6	2	2	2	4	2	1	1	1	1	1	1	1	1
7	3	2	4	1	4	1	3	3	3	1	2	2	1
8	4	2	1	3	1	3	4	1	3	1	2	1	2
9	1	3	1	4	4	2	4	3	2	1	1	1	1
10	2	3	4	2	1	4	3	1	2	1	1	2	2
11	3	3	2	3	3	4	1	3	4	1	2	1	2
12	4	3	3	1	2	2	2	1	4	1	2	2	1
13	1	4	2	1	1	4	2	4	3	1	1	1	1
14	2	4	3	3	4	2	1	2	3	1	1	2	2
15	3	4	1	2	2	2	3	4	1	1	2	1	2
16	4	4	4	4	3	4	4	2	1	1	2	2	1
17	1	1	4	2	3	2	1	1	3	2	2	1	1
18	2	1	1	4	2	4	2	3	3	2	2	2	2
19	3	1	3	1	4	4	4	1	1	2	1	1	2
20	4	1	2	3	1	2	3	3	1	2	1	2	1
21	1	2	3	3	2	4	3	2	2	2	2	1	1
22	2	2	2	1	3	2	4	4	2	2	2	2	2
23	3	2	4	4	1	2	2	2	4	2	1	1	2
24	4	2	1	2	4	4	1	4	4	2	1	2	1
25	1	3	1	1	1	1	1	2	1	2	2	2	2
26	2	3	4	3	4	3	2	4	1	2	2	1	1
27	3	3	2	2	2	3	4	2	3	2	1	2	1
28	4	3	3	4	3	1	3	4	3	2	1	1	2
29	1	4	2	4	4	3	3	1	4	2	2	2	2
30	2	4	3	2	1	1	4	3	4	2	2	1	1
31	3	4	1	3	3	1	2	1	2	2	1	2	1
32	4	4	4	1	2	3	1	3	2	2	1	1	2

$$PB_{54}(3^{25} \times 2^1)$$

条件号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	1	2	0	1	2	1	2	0	1	2	1	2	0	1	2	1	2	0	1	2
3	0	2	0	0	0	0	2	1	1	0	2	2	1	1	0	2	2	1	1	0	2	2	1	1	0	2
4	1	0	0	0	0	0	0	2	1	2	1	0	2	1	2	1	0	2	1	2	1	0	2	1	2	1
5	1	1	0	0	0	0	2	0	2	1	1	2	0	2	1	1	2	0	2	1	1	2	0	2	1	1
6	1	2	0	0	0	0	1	1	2	2	0	1	1	2	2	0	1	1	2	2	0	1	1	2	2	0
7	0	0	0	1	1	2	0	0	0	0	0	1	1	1	1	1	2	2	2	2	2	1	1	1	1	1
8	0	1	0	1	1	2	1	2	0	1	2	2	0	1	2	0	0	1	2	0	1	2	0	1	2	0
9	0	2	0	1	1	2	2	1	1	0	2	0	2	2	1	0	1	0	0	2	1	0	2	2	1	0
10	1	0	0	1	1	2	0	2	1	2	1	1	0	2	0	2	2	1	0	1	0	1	0	2	0	2
11	1	1	0	1	1	2	2	0	2	1	1	0	1	0	2	2	1	2	1	0	0	0	1	0	2	2
12	1	2	0	1	1	2	1	1	2	2	0	2	2	0	0	1	0	0	1	1	2	2	2	0	0	1
13	0	0	0	2	2	1	0	0	0	0	0	2	2	2	2	2	1	1	1	1	1	2	2	2	2	2
14	0	1	0	2	2	1	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	0	1	2	0	1
15	0	2	0	2	2	1	2	1	1	0	2	1	0	0	2	1	0	2	2	1	0	1	0	0	2	1
16	1	0	0	2	2	1	0	2	1	2	0	2	1	0	1	0	1	0	2	0	2	2	1	0	1	0
17	1	1	0	2	2	1	2	0	2	1	1	1	2	1	0	0	0	1	0	2	2	1	2	1	0	0
18	1	2	1	2	2	1	1	1	2	2	0	0	0	1	1	2	2	2	0	0	1	0	0	1	1	2
19	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
20	0	1	1	0	1	1	2	0	1	2	0	2	0	1	2	0	2	0	1	2	0	1	2	0	1	2
21	0	2	1	0	1	1	0	2	2	1	0	0	2	2	1	0	0	2	2	1	0	2	1	1	0	2
22	1	0	1	0	1	1	1	0	2	0	2	1	0	2	0	2	1	0	2	0	2	0	2	1	2	1
23	1	1	1	0	1	1	0	1	0	2	2	0	1	0	2	2	0	1	0	2	2	2	0	2	1	1
24	1	2	1	0	1	1	2	2	0	0	1	2	2	0	0	1	2	2	0	0	1	1	1	2	2	0
25	0	0	1	1	2	0	1	1	1	1	1	2	2	2	2	2	0	0	0	0	0	1	1	1	1	1
26	0	1	1	1	2	0	2	0	1	2	0	0	1	2	0	1	1	2	0	1	2	2	0	1	2	0
27	0	2	1	1	2	0	0	2	2	1	0	1	0	0	2	1	2	1	1	0	2	0	2	2	1	0
28	1	0	1	1	2	0	1	0	2	0	2	2	1	0	1	0	0	2	1	2	1	1	0	2	0	2
29	1	1	1	1	2	0	0	1	0	2	2	1	2	1	0	0	2	0	2	1	1	0	1	0	2	2
30	1	2	1	1	2	0	2	2	0	0	1	0	0	1	1	2	1	1	2	2	0	2	2	0	0	1
31	0	0	1	2	0	2	1	1	1	1	1	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2
32	0	1	1	2	0	2	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	0	1	2	0	1
33	0	2	1	2	0	2	0	2	2	1	0	2	1	1	0	2	1	0	0	2	1	1	0	0	2	1
34	1	0	1	2	0	2	1	0	2	0	2	0	2	1	2	1	2	1	0	1	0	2	1	0	1	0
35	1	1	1	2	0	2	0	1	0	2	2	2	0	2	1	1	1	2	1	0	0	1	2	1	0	0
36	1	2	2	2	0	2	2	2	0	0	1	1	1	2	2	0	0	0	1	1	2	0	0	1	1	2
37	0	0	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0
38	0	1	2	0	2	2	0	1	2	0	1	0	1	2	0	1	0	1	2	0	1	1	2	0	1	2
39	0	2	2	0	2	2	1	0	0	2	1	1	0	0	2	1	1	0	0	2	1	2	1	1	0	2
40	1	0	2	0	2	2	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	2	1	2	1

附录 B: 实验用正交表及 PB 表

41	1	1	2	0	2	2	1	2	1	0	0	1	2	1	0	0	1	2	1	0	0	2	0	2	1	1
42	1	2	2	0	2	2	0	0	1	1	2	0	0	1	1	2	0	0	1	1	2	1	1	2	2	0
43	0	0	2	1	0	1	2	2	2	2	2	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
44	0	1	2	1	0	1	0	1	2	0	1	1	2	0	1	2	2	0	1	2	0	2	0	1	2	0
45	0	2	2	1	0	1	1	0	0	2	1	2	1	1	0	2	0	2	2	1	0	0	2	2	1	0
46	1	0	2	1	0	1	2	1	0	1	0	0	2	1	2	1	1	0	2	0	2	1	0	2	0	2
47	1	1	2	1	0	1	1	2	1	0	0	2	0	2	1	1	0	1	0	2	2	0	1	0	2	2
48	1	2	2	1	0	1	0	0	1	1	2	1	1	2	2	0	2	2	0	0	1	2	2	0	0	1
49	0	0	2	2	1	0	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0	2	2	2	2	2
50	0	1	2	2	1	0	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1
51	0	2	2	2	1	0	1	0	0	2	1	0	2	2	1	0	2	1	1	0	2	1	0	0	2	1
52	1	0	2	2	1	0	2	1	0	1	0	1	0	2	0	2	0	2	1	2	1	2	1	0	1	0
53	1	1	2	2	1	0	1	2	1	0	0	0	1	0	2	2	2	0	2	1	1	1	2	1	0	0
54	1	2	2	2	1	0	0	0	1	1	2	2	2	0	0	1	1	1	2	2	0	0	0	1	1	2

$$PB_{48}(2^{11} \times 4^{12})$$

条件号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	3	3	3	1	1	1	2	2	2	0	0	0	0	0	1	1	1	1	1	1
3	0	0	0	1	1	1	2	2	2	3	3	3	0	0	1	1	1	1	1	1	0	0	0
4	0	1	3	2	3	1	3	2	0	1	0	2	1	0	1	0	1	0	1	0	1	0	1
5	0	1	3	1	2	3	0	3	2	2	1	0	1	0	1	1	0	0	0	1	1	1	0
6	0	1	3	3	1	2	2	0	3	0	2	1	1	0	0	1	1	1	0	0	0	1	1
7	0	3	2	1	0	2	1	3	0	1	2	3	0	1	1	0	1	0	0	1	0	1	1
8	0	3	2	2	1	0	0	1	3	3	1	2	0	1	1	1	0	1	0	0	1	0	1
9	0	3	2	0	2	1	3	0	1	2	3	1	0	1	0	1	1	0	1	0	1	1	0
10	0	2	1	3	2	0	1	2	3	1	3	0	1	1	0	1	0	0	1	1	0	0	1
11	0	2	1	0	3	2	3	1	2	0	1	3	1	1	0	0	1	1	0	1	1	0	0
12	0	2	1	2	0	3	2	3	1	3	0	1	1	1	1	0	0	1	1	0	0	1	0
13	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
14	1	1	1	2	2	2	0	0	0	3	3	3	0	0	0	0	0	1	1	1	1	1	1
15	1	1	1	0	0	0	3	3	3	2	2	2	0	0	1	1	1	1	1	1	0	0	0
16	1	0	2	3	2	0	2	3	1	0	1	3	1	0	1	0	1	0	1	0	1	0	1
17	1	0	2	0	3	2	1	2	3	3	0	1	1	0	1	1	0	0	0	1	1	1	0
18	1	0	2	2	0	3	3	1	2	1	3	0	1	0	0	1	1	1	0	0	0	1	1
19	1	2	3	0	1	3	0	2	1	0	3	2	0	1	1	0	1	0	0	1	0	1	1
20	1	2	3	3	0	1	1	0	2	2	0	3	0	1	1	1	0	1	0	0	1	0	1
21	1	2	3	1	3	0	2	1	0	3	2	0	0	1	0	1	1	0	1	0	1	1	0

22	1	3	0	2	3	1	0	3	2	0	2	1	1	1	0	1	0	0	1	1	0	0	1
23	1	3	0	1	2	3	2	0	3	1	0	2	1	1	0	0	1	1	0	1	1	0	0
24	1	3	0	3	1	2	3	2	0	2	1	0	1	1	1	0	0	1	1	0	0	1	0
25	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
26	2	2	2	1	1	1	3	3	3	0	0	0	0	0	0	0	0	1	1	1	1	1	1
27	2	2	2	3	3	3	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0	0	0
28	2	3	1	0	1	3	1	0	2	3	2	0	1	0	1	0	1	0	1	0	1	0	1
29	2	3	1	3	0	1	2	1	0	0	3	2	1	0	1	1	0	0	0	1	1	1	0
30	2	3	1	1	3	0	0	2	1	2	0	3	1	0	0	1	1	1	0	0	0	1	1
31	2	1	0	3	2	0	3	1	2	3	0	1	0	1	1	0	1	0	0	1	0	1	1
32	2	1	0	0	3	2	2	3	1	1	3	0	0	1	1	1	0	1	0	0	1	0	1
33	2	1	0	2	0	3	1	2	3	0	1	3	0	1	0	1	1	0	1	0	1	1	0
34	2	0	3	1	0	2	3	0	1	3	1	2	1	1	0	1	0	0	1	1	0	0	1
35	2	0	3	2	1	0	1	3	0	2	3	1	1	1	0	0	1	1	0	1	1	0	0
36	2	0	3	0	2	1	0	1	3	1	2	3	1	1	1	0	0	1	1	0	0	1	0
37	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
38	3	3	3	0	0	0	2	2	2	1	1	1	0	0	0	0	0	1	1	1	1	1	1
39	3	3	3	2	2	2	1	1	1	0	0	0	0	0	1	1	1	1	1	1	0	0	0
40	3	2	0	1	0	2	0	1	3	2	3	1	1	0	1	0	1	0	1	0	1	0	1
41	3	2	0	2	1	0	3	0	1	1	2	3	1	0	1	1	0	0	0	1	1	1	0
42	3	2	0	0	2	1	1	3	0	3	1	2	1	0	0	1	1	1	0	0	0	1	1
43	3	0	1	2	3	1	2	0	3	2	1	0	0	1	1	0	1	0	0	1	0	1	1
44	3	0	1	1	2	3	3	2	0	0	2	1	0	1	1	1	0	1	0	0	1	0	1
45	3	0	1	3	1	2	0	3	2	1	0	2	0	1	0	1	1	0	1	0	1	1	0
46	3	1	2	0	1	3	2	1	0	2	0	3	1	1	0	1	0	0	1	1	0	0	1
47	3	1	2	3	0	1	0	2	1	3	2	0	1	1	0	0	1	1	0	1	1	0	0
48	3	1	2	1	3	0	1	0	2	0	3	2	1	1	1	0	0	1	1	0	0	1	0

发表文章目录

基金项目:

- [1] 汉语框架语义角色自动标注技术研究,国家自然科学基金(No.60873128)

发表文章:

- [1] 王瑞波, 李国臣, 李济洪. 条件随机场模型归一化因子的计算方法. 山西大学学报增刊, 2009, **32**(1):17-20
- [2] 李国臣, 王瑞波, 李济洪. 基于条件随机场模型的汉语功能块自动标注. 计算机研究与发展 (已录用)
- [3] 李济洪, 王瑞波, 王凯华, 李国臣. 基于最大熵模型的中文阅读理解问题回答技术研究. 中文信息学报, 2008, **22**(2):55-62
- [4] 李济洪, 王瑞波, 王蔚林, 杨杏丽, 高亚慧. 汉语框架语义角色的自动标注研究进展. 全国第十届计算语言学学术会议(CNCCL-2009), 2009 (已录用)
- [5] 李济洪, 杨杏丽, 王瑞波, 张娜, 李国臣. 基于规则的中文阅读理解问题回答技术研究. 中文信息学报 (已录用)
- [6] 王凯华, 李济洪, 张国华, 王瑞波. 基于最大熵模型的中文阅读理解问答系统技术研究, 第九届全国计算语言学学术会议论文集, 清华大学出版社, 2007, 643-648
- [7] 李国臣, 王瑞波, 李济洪. 基于基本块的汉语功能块自动标注, 第四届学生计算语言学研讨会, 2008
- [8] 李济洪, 王凯华, 王瑞波. 基于最大熵模型的中文阅读理解技术研究, 第四届学生计算语言学研讨会, 2008

致 谢

值此论文完成之际，谨向给予我无私帮助的老师 and 同学们致以诚挚的谢意！

首先感谢我的导师李国臣教授，我的每一步成长跟他的教诲是分不开的。我在山西大学研究生近三年的学习和研究中倾注了李老师的一片心血。李老师不仅为我提供了极其优越的实验环境，营造了轻松和谐的学习氛围，而且给我创造了学术交流的条件和机会，开阔了我的眼界。李老师渊博的专业知识、敏锐的学术洞察力、严谨的治学态度，忘我的工作精神，对事业孜孜不倦的追求和创新精神给我留下了难忘的记忆，让我终生受益，并将永远激励我奋发进取。

尤其感谢李济洪教授和李茹教授。本文从选题、研究到最终定稿都得到了两位教授的悉心指导和帮助。另外，两位教授组织的讨论课让我分别从理论和应用两个不同的角度来了解汉语框架网络的意义和作用，为我今后的研究奠定了重要的基础。

三年来，计算机与信息技术学院以及语义网研究室的老师和同学们在学习、生活上给予了我大量的指导和帮助，使我得以顺利地完成硕士研究生阶段的学习。

本文采用的汉语框架网络知识库的语料资源是由刘开瑛教授、由丽萍博士等老师和同学们提供的，并且本文所使用的汉语基本块分析器是由清华大学周强教授提供的，在此对他们表示衷心地感谢！

感谢我的师哥师姐王凯华、张国华、尹晓丽和张娜对我学习和生活上的关心和照顾。虽然她们现在已经在各自的岗位上辛勤地工作了一年多，但和他们共同学习生活的场景至今历历在目。

感谢和我一起进行课题研究的王蔚林同学，杨杏丽同学和高亚慧同学。我们一起进行讨论和研究，使我受益匪浅。特别感谢高亚慧同学对本文的认真审阅和校对。

感谢所有关心并支持我的亲人和朋友，由衷地感谢我的父母一直以来的无私奉献，他们给了我巨大的支持和鼓励。

最后衷心感谢各位老师百忙中抽出宝贵时间评审我的论文并参加我的硕士论文答辩会。

个人简历

个人概况：

王瑞波（1985.8-），男，山西省长治市潞城人。

个人简况：

2006.9-2009.7 山西大学 计算机与信息技术学院 软件与理论专业 硕士研究生
研究方向：统计自然语言处理

2002.9-2006.7 山西师范大学 数学与计算机科学学院 信息与计算科学专业 本科

联系方式：

电子信箱：wangruiibo@sxu.edu.cn

承 诺 书

本人郑重声明：所呈交的学位论文，是在导师指导下独立完成的，学位论文的知识产权属于山西大学。如果今后以其他单位名义发表与在读期间学位论文相关的内容，将承担法律责任。除文中已经注明引用的文献资料外，本学位论文不包括任何其他个人或集体已经发表或撰写过的成果。

学位论文作者（签章）：

200 年 月 日