

# 语义角色在指代消解中的研究\*

王海东 谭魏旋 周国栋

(苏州大学计算机科学与技术学院, 江苏, 苏州, 215006)

(江苏省计算机信息处理技术重点实验室, 江苏, 苏州, 215006)

{064227065056, 20074227162092, gdzhou}@suda.edu.cn

**摘要:** 本文实现了一个基于机器学习的指代消解平台。在此基础上, 通过自动语义角色标注工具得到目标动词和语义角色信息, 组合语义角色特征和动词驱动特征, 研究发现两者的结合能够显著提高系统的性能。对于无法得到语义类别的名词, 本文尝试利用动词的选择限制来给这类名词赋予语义类别。在对 ACE2003 NWIRE 中所有类型名词短语的测试表明, 综合考虑语义角色、动词驱动以及动词的选择限制能显著提高系统的召回率和 F 值。

**关键字:** 指代消解, 语义角色标注, 动词驱动特征, 动词选择限制

## Semantic Role Labeling for Coreference Resolution

Haidong Wang WeiXuan Tan Guodong Zhou

(School of Computer Science & Technology, Soochow University, Suzhou, China, 215006)

(Jiangsu Provincial Key Lab for Computer Information Processing Technology, Suzhou, China, 215006)

{064227065056, 20074227162092, gdzhou}@suda.edu.cn

**Abstract:** This paper proposes a machine learning-based platform to coreference resolution, in which various semantic role-based and predicate-driven features are incorporated. Here, semantic roles are acquired automatically from a state-of-the-art semantic role labeling toolkit. In addition, for those noun phrases hard to determine their semantic classes, we attempt to determine their classes using the selectional preference of the corresponding predicate. Evaluation on the ACE-2003 NWIRE benchmark corpus shows that the incorporation of semantic role information, predicate-driven features and the selectional preference can much improve the performance, especially in recall and F-measure.

**Keywords:** Coreference Resolution, Semantic Roles Labeling, Predicate-driven Feature, Selectional Preference

## 1 引言

指代是自然语言处理的关键问题之一, 随着篇章处理的相关应用日益广泛, 指代消解也显示出前所未有的重要性, 指代消解是自然语言处理中文本理解不可缺少的内容, 在文本摘要、机器翻译、多语言信息处理和信息抽取等诸多应用中都涉及到指代消解问题。一般情况下, 指代分为两种, 回指(Anaphora, 也称为指示性指代) 和共指(Coreference, 也称同指)。回指是指当前的指示词与上文出现的词、短语或句子存在密切的关系; 共指则主要是指两个名词短语指向真实世界中的同一个实体<sup>[1]</sup>。

\* 本文受国家自然科学基金项目(60673041)和863高技术项目(2006AA01Z147)资助

作者简介: 王海东(1981—), 男, 硕士研究生, 研究方向自然语言处理; 谭魏旋(1986—), 女, 硕士研究生, 研究方向自然语言处理; 周国栋(1967—), 男, 教授, 博士生导师, 研究方向自然语言处理

早期指代消解研究主要利用了领域和语言语法知识形成逻辑规则进行消解,比较具有代表性的方法包括:Hobbs算法,中心理论,基于句法的方法等。此类方法的主要问题是表示和处理都非常困难,且需要大量的人工输入,同时系统可移植性和自动化程度也比较差。近年来,有监督的机器学习方法在指代消解中得到了广泛应用。Soon(2001)<sup>[2]</sup>采用了基于决策树的机器学习方法,抽取了12个基本特征,并给出了完整步骤和实现平台,在MUC6上达到了F值62.6,该方法也成为了国内外利用机器学习方法进行指代消解研究的基础。Ng等(2002)<sup>[3]</sup>对Soon的研究进行了扩充,提取了53个特征,在MUC6上F值达到了69.4。后来的大多数研究则是针对某种类别的名词短语进行消解,增加先行语和照应语的语义信息和结构化特征。例如Yang等(2004)<sup>[4]</sup>、(2005)<sup>[5]</sup>、(2006)<sup>[6]</sup>分别从先行语的指代关系、语义和结构化信息三个方面对代名词消解进行改进。Yang等(2004)<sup>[3]</sup>提出了使用了指代链的特征并在其中考虑了名词短语作为主语或宾语的情况,但是他并没有单独分析语义角色的特点。SP Ponzetto等(2006)<sup>[7]</sup>考虑了语义角色特征,但是仅仅考虑了照应语和先行语是否做主语和宾语角色,并没有用到语义角色标注的其它信息,特别是没有考虑动词的选择限制对指代消解系统的影响。

本文采用机器学习的方法,在Soon的基础上进行扩展,实现了一个指代消解平台,重点讨论了名词短语的语义角色特征、动词驱动特征以及动词的选择限制对指代消解的影响。在ACE2003 NWIRE 基准语料库上的实验表明,语义角色特征和动词的选择限制能显著提高指代消解系统的性能,相对它们加入前召回率提高了1.7%,F值提高了0.9%,其中代词的F值更是提高了2.8%。

## 2 系统原型

本系统采用基于语料库的机器学习方法进行指代消解,使用SVMlight<sup>1</sup>作为分类器。原型系统的特征使用Soon等(2002)的12个基本特征。原始文本进入系统之前,要进行如下预处理:句子边界识别、分词、词性标注、命名实体识别、语义类别的获取、句法分析等。本文使用Zhou等(2004)<sup>[8]</sup>的预处理系统,其中命名实体识别、词性标注以及名词短语识别都是基于错误驱动的HMM模型(Zhou and Su 2000);名词短语的中心词获得用的是Collins的方法;语义信息主要从命名实体识别程序和WordNet中获取。

在训练阶段,预处理前我们先将指代链信息抽取到一个文件中,过滤掉训练文件中的指代信息后再进行预处理,预处理完成后再把处理生成的名词短语和我们读取的指代信息进行匹配,然后根据名词短语的指代信息生成正负例进行训练。在测试阶段,先过滤掉测试语料中指代信息,然后进行预处理。预处理后将根据选取的特征生成测试用例交给SVM进行测试,从而确定当前照应语和先行词是否有指代关系。

## 3 语义角色特征、动词特征及动词的选择限制

浅层语义分析是目前自然语言处理的一个热点问题。作为一个可行的解决方法,近几年人们对语义角色标注作了广泛的研究。对英文浅层语义分析,国际上分别于2004和2005年举行过三

---

<sup>1</sup> <http://svmlight.joachims.org/>

次评测，分别为 Senserval-3 以及 CoNLL 主办的 SRL Share Task 2004 和 2005。

### 3.1 语义角色特征

语义角色是句子中的名词短语在相应动词驱动下所承担的句法成分。语义角色的获得目前所用的主要方法是语义角色标注 (Semantic Role Labeling)。语义角色标注是浅层语义分析 (Shallow Semantic Parsing) 的一种实现方式。所谓浅层语义分析，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的某种形式化表示。该方法并不对整个句子进行详细的语义分析，而仅仅标注句子中的一些成份为给定谓词的语义角色。核心的语义角色为 Arg0-5 六种，Arg0 通常表示动作的施事，Arg1 通常表示动作的影响等。其余的语义角色为附加语义角色，使用 ArgM 表示，如 ArgM-LOC 表示地点，ArgM-TMP 表示时间等。另外对于驱动动词使用 TARGET 表示。例如 “[委员会 Arg0][明天 ArgM-TMP]将要[通过 TARGET][此议案 Arg1]。”中，“通过”为谓词，“委员会”、“此议案”和“明天”分别是其施事、受事和发生的时间。

本文使用 ASSERT<sup>2</sup>语义角色标注系统得到语义角色标注信息。ASSERT 是通过动词驱动的角色标注。对于同一个句子可能出现多个动词，那么对每一个动词分别进行语义角色标注。由中心理论<sup>9,10</sup>可知，不同语法成份担当语义角色时有一定的优先级。一般表现为：主语 (subject) > 宾语 (object) > 间接宾语 > 其他名词，由于其它名词在句子中没有突显性，所以我们这里暂不考虑其他名词，只考虑主语和宾语成份，也就是 Arg0 和 Arg1。

由于 ASSERT 是目标动词驱动的角色标注，而一般英语句子通常不止一个动词，如图 1 所示，该句子有两个目标动词。一个名词短语可能作为某个动词的主语，同时又作为另一个动词的宾语。在本文中，暂不考虑名词短语被多个动词驱动，只关心它是否被某些动词驱动作主语或宾语。例如，如果一个名词短语被两个动词驱动，都是作主语，那么就记它作主语，如果被某个动词驱动作主语，被另一个动词驱动作宾语，那么记它作主语也作宾语。所以，本文为照应语和先行词分别设置两个特征，一个表示其在句子中是否被某个动词驱动作主语，另一个表示其在句子中是否被某个动词驱动作为宾语，如表 2 所示。

表 2: 语义角色标注特征

语义角色特征	
ANArg0	照应语在句子中被某个动词驱动作主语取 1，否则取 0
ANArg1	照应语在句子中被某个动词驱动作宾语取 1，否则取 0
CAArg0	先行语在句子中被某个动词驱动作主语取 1，否则取 0
CAArg1	先行语在句子中被某个动词驱动作宾语取 1，否则取 0
动词特征	
ANArgVerb	照应语在句子中被某个动词驱动，并且该动词是句子的主动词时取 1，否则取 0
CAArgVerb	先行语在句子中被某个动词驱动，并且该动词是句子的主动词时取 1，否则取 0
ANCABothSentTarget	先行语和照应语在同一个句子中且被同一个动词驱动时取 1，否则取 0

<sup>2</sup> assert-v0.14b Sameer Pradhan and Steven Bethard

### 3.2 动词特征

每一个句子中都有一个主动词，同时可能还有其它的从句动词。在图 1 中，该语句有两个目标动词，[demanded]是主动词，[broke]是由 but 引导的从句中的动词。对于主动词驱动的名词短语要比其它动词驱动的名词短语有更高的优先级。本文考虑了驱动名词短语的动词是不是主动词。这里增加了两个特征，一个表示照应语是否被主动词驱动，另一个表示先行语是否被主动词驱动。如果两个词被同一个动词驱动，往往两者之间是不可能指代关系。所以我们同时增加了是否被同一个动词驱动这样的特征。具体的语义角色标注特征见表 2 所示。

图 1: 通过 ASSERT 标注的 ACE2003 NWIRE 的片段

```
4: [ARGO His kidnapers] [TARGET demanded] [ARG1 a ransom of dlrs 1 million] , but negotiations broke off at noon Thursday .
4: His kidnapers demanded a ransom of dlrs 1 million , but [ARG1 negotiations] [TARGET broke] off [ARGM-TMP at noon Thursday] .
```

### 3.3 动词的选择限制

“It”一直是指代消解的一个难点，因为它包含的信息很少，而且它除了不指代人外，可以指代任何名词短语；另外在我们系统的预处理过程中，经常遇到一些未登录词，因为在 WordNet 中查不到这个词，我们就将其表示为 bareNp，并认为它和所有的名词短语都是语义类别一致的，这无疑引入了很多噪音。通过动词对它的论元所施加的选择限制可以帮助对上面那类词的指代消解。如图 2 中所示，Fernandez 是一个未登录词，我们无法知道它的语义类别，但是通过动词 handed，就可以知道作为动词 hand 施事的名词一般属于 person。见下面的图 3。所以系统就可以将 Fernandez 划到 person 类别中去，这样有助于指代消解。

图 2: 选择限制的主要作用

```
<s> Last year, Fernandez handed over to U.S. authorities two suspects accused in U.S. courts of multiple murders, drug trafficking and conspiracy. </s>
```

由于英文中动词很多，而且每一个动词所驱动的语义类别也不一样，目前还没有见到关于动词选择限制的相关研究。由于暂时无法得到所有动词驱动的语义类别，所以在本系统中只是通过 Assert 得到训练文件中出现的动词，然后通过 WordNet 找出动词驱动的名词短语所属的语义类别，最后得到动词驱动的语义类别。由于动词对于主语和宾语的驱动有不一样的形式，所以在 这里我们分别考虑这两种情况。对于语义类别，我们使用 WordNet 中的 24 类。由于训练语料和测试语料同属于一类，因此训练语料中出现的大部分动词在训练语料中也很可能出现。所以我们可以 通过训练语料得到部分动词的选择限制，然后去解决测试语料中出现的没有语义类别的名词。这样能部分解决无法取得所有动词的选择限制的问题。

我们将从训练语料得到的动词及其所驱动的名词所属的语义类别组织成一个文件，见图 3 所示。图中是一个动词的选择限制，前面三个词是这个动词的时态变化形式，后面是可以做该动词主语的名词语义类。在测试过程中，如果当前的名词短语在句子中作为主语或宾语，并且不能获得该名词短语的语义类别信息，那么我们就可以得到驱动该名词短语的动词，然后将该动词所驱动的名词短语的语义类别赋予该名词短语。

图 3: 动词选择限制的一个示例

```
hand handing handed |PERSON|
```

## 4 实验结果与分析

为了分析加入动词特征对系统性能的影响,我们将语义角色标注特征分为两类:语义角色特征和动词特征。见表 2 所示。我们使用 ACE2003 NWIRE 的语料作为训练和测试集,并采用 MUC 的标准评测程序进行评测,实验结果如表 3 所示。

表 3: 对所有名词短语,在 ACE2003 语料上的测试结果

System Variation	R	P	F
Baseline: Our Duplicated Soon 等 (2001) .	51.9	64.5	57.5
Ng (2005): Perfect ranking model	56.6	59.7	58.1
Baseline + 语义角色特征	53.4	62.7	57.7
Baseline + 语义角色特征 + 动词驱动特征	<b>53.6</b>	<b>64.1</b>	<b>58.4</b>

### 4.1 语义角色特征和动词驱动特征的贡献分析

目前在 ACE2003 上的相关研究比较少,本文列出了 Ng (2005) 在相同语料上的测试结果,于 Ng 的结果相比,本文系统的性能要稍好于 Ng 的系统,准确率比 Ng 的系统高,而召回率稍有下降。

从实验结果上看,在原型系统上加入语义角色特征能明显提高系统的召回率,在 NWIRE 测试语料上提升了 1.5%,而准确率有所降低。在加入动词特征后,系统的召回率没有什么变化,但是准确率得到了明显的提升,准确率提升了 1.8%。

通过对实验得到的指代结果分析得知,语义角色特征对一个句子内有固定结构的代词有较好的消解结果,因为一般的代词在句子中都作为语义角色,主要是主语或宾语。所以通过加入语义角色特征,增加了代词的权重,使代词得到了很好的消解。虽然使召回率得到了提升,但因为不区分 Arg0 和 Arg1 之间谁更突显,所以同样也带进来很多的噪音。比如在同一个句子内,同一个动词驱动的主语和宾语很少有指代关系;同时,主动词驱动的语义角色要比其他的动词驱动的语义角色有更大的权重,所以在加入动词特征后,很好的解决了上述问题,系统的准确率得到了较大的提升。

表 4: 仅对代名词,在 ACE2003 语料上的测试结果

System Variation	R	P	F
Baseline: Duplicated Soon 等 (2001) .	43.0	68.2	52.7
Baseline + 语义角色特征	46.9	62.7	53.7
Baseline + 语义角色特征 + 动词特征	<b>47.0</b>	<b>67.8</b>	<b>55.5</b>

对消解的结果进行分析,我们发现语义角色信息对代词的消解有很好的效果,所以我们单独考虑代词的消解,表 4 是代词的消解结果。从表中所见,在加入语义角色特征和动词特征后,系统的召回率提升了 4%,而准确率没有太大的变化。

表 5: 对所有名词短语,在 ACE2003 语料上的测试结果

System Variation	R	P	F
Baseline + 语义角色标注特征	53.6	64.1	58.4
Baseline + 语义角色标注特征 + 动词选择限制	<b>53.3</b>	<b>64.5</b>	<b>58.5</b>

## 4.2 动词选择限制的结果分析

从表 5 看, 在加了动词的选择限制以后, 系统的准确率提升了 0.4%, 但同时召回率也有一定的下降。这里说明动词的选择限制还是对系统的准确消解起到一定的作用, 但同时也引入了部分噪音。虽然系统的总体性能没有太大的改变, 但并不能说明动词的选择限制是没有用的。分析主要有以下几点原因使得加入动词限制后系统的性能没有变化:

1. 在测试语料中, 对于动词的匹配成功的比例太小。

在 ACE2003 NWIRE 的测试语料中, 没有语义类别的名词主要是 it 和未登录词, 共有 378 处未登录词和 271 处 “it”, 而通过动词匹配, 得到语义类别的只有 89 处, 所以得到的驱动名词语义类别的动词很稀疏。并且由于动词存在时态的变化, 也造成了一定数量的匹配错误。

2. 名词通常有多种语义, 这也造成了动词驱动的语义类别不够准确。

如表 1 所示, 在本文的指代消解平台中, 虽然使用 soon 等 (2001) 的 12 个基本特征, 但语义类别特征不同于 soon 的语义类别。本文的语义类别采用 WordNet 的名词分类, 共分为 24 类。在预处理的过程中, 进行命名实体的识别, 对于命名实体, 它的语义类别就是指命名实体的类别。命名实体识别程序没有识别出来的名词短语, 本文首先获得它的中心词, 然后在 WordNet 中查找其语义类别信息。通过 WordNet 得到的语义类别往往有多个, 而在确定的上下文中, 一个名词短语只应有一个语义类别, 我们目前并不对其进行消歧, 所以对于两个名词短语, 只要它们有一个义项是属于同一个类别, 那么我们就认为他们语义类别一致。这样也就削弱了语义类别一致这个特征的作用, 同样也就削弱了动词限制的作用。

## 5 结论

本文重点讨论了语义角色特征和动词驱动特征对指代消解系统的影响, 主要从照应语和先行语的语义角色以及该语义角色是否被主动词驱动出发, 详细讨论了它们对系统性能的贡献。在 ACE2003 NWIRE 基准语料上的测试表明, 加入语义角色特征和动词特征可以显著的提高系统召回率和 F 值, 其中召回率提升了 1.7%, F 值提升了 0.9%。

在以上的基础上, 本文还对动词选择限制如何用于指代消解进行了一些初步尝试。实验表明, 加入动词的选择限制后, 系统的准确率得到了一定的提升, 召回率有所下降。通过上一节的原因分析, 本文的下一步工作主要应是以下几点:

1. 在目前的实验中, 只是简单的得到训练语料中的动词, 然后将这些动词运用于测试语料, 实验效果并不理想。下一步的工作是: 如何得到更多动词的选择限制, 使在测试的过程中能够尽量匹配所有的动词。
2. 在本文的指代消解平台中, 一个名词往往因为有多个义项而使语义类别的判断不够准确, 下一步可以尝试用一些语义消歧工具对语义类别进行消歧, 从而更好的发挥动词的选择限制作用。

图 4: 动词选择限制的作用<sup>3</sup>

John parked his *Acura* in the *garage*. He had driven it around for hours.

3. 还需要进一步考虑名词划分粒度。如果名词类别划分的不够细,就不能很好的体现各种动词的选择限制。如图 4 所示, it 有两个可能的所指对象, Acura 和 garage, 而动词 drive 要求它的直接宾语是某种能够驾驶的事物, 比如汽车 (car)、公共汽车 (bus) 等, 但不能是车库 (garage)。本文用目前的 WordNet 的 24 类, Acura 和 garage 都属于 object, 并不能区分两者; 但如果使用更细的粒度 vehicle (交通工具), 就可以根据动词的选择限制很好的区分两者。如要将名词类别划分的足够细, 就需要大量的训练语料来得到动词的选择限制, 这一点目前没有很好的办法。所以如何在现有的基础上选择一个合适的名词划分粒度也是本文下面的研究工作。

### 参考文献

- [1]. 王厚峰. 指代消解的基本方法和实现技术[J]. 中文信息学报, 2002, (6): 9-17.
- [2]. Wee Meng Soon, Hwee Tou Ng and Daniel Chung Yong Lim. A machine learning approach to coreference resolution of noun phrase. *Computational Linguistics*, 2001, 27(4): 521-544.
- [3]. Vincent Ng and Claire Cardie. Improving machine learning approaches to coreference resolution. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002.
- [4]. Yang X.F., Su J., Zhou G.D. and Tan C.L. 2004. Improving pronoun resolution by incorporating coreferential information of candidates. *ACL' 2004*: 127-134, Barcelona, Spain, 21-26 July 2004.
- [5]. Yang X.F., Zhou G.D., Su J. and Tan C.L. 2003. Coreference Resolution Using Competition Learning Approach, *ACL' 2003*: 176-183, Sapporo, Japan, 7-12 July 2003.
- [6]. Yang X.F., Su J. and Tan C.L. 2006. Kernel-Based Pronoun Resolution with Structured Syntactic Knowledge. *ACL' 2006*: 41-48, Sydney, July 2006.
- [7]. Simone Paolo Ponzetto and Michael Strube. 2006. Semantic Role Labeling for Coreference Resolution, *Companion Volume of the Proceedings of the 11th Meeting*, 2006.
- [8]. Zhou G.D. and Su J. 2004. A high-performance coreference resolution system using a multi-agent strategy, *COLING' 2004*: 522-528, 23-27 Aug, 2004, Geneva, Switzerland.
- [9]. Grosz, A. JoShi, and S. Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual meeting of the Association for Computational Linguistics*, pages 44-45.
- [10]. Grosz, A. JoShi and S. Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2): 203-225.
- [11]. Vincent Ng. Semantic Class Induction and Coreference Resolution *ACL' 2007* 536-543 Prague, Czech Republic, June 2007.
- [12]. Yang X.F., Su J. and Tan C.L. 2005. Improving Pronoun Resolution Using Statistics - Based Semantic Compatibility Information, *ACL' 2005*: 165-172, Univ of Michigan-Ann Arbor, USA, 25-30 June 2005
- [13]. Ng, V. (2007). Shallow semantics for coreference resolution. In *IJCAI 2007*, pages 1689-1694.

<sup>3</sup> 本示例来自于《自然语言处理综论》, Daniel Jurafsky and James H. Martin 著, 2005-6, 电子工业出版社。