

语义角色标注的方法与挑战

车万翔, 刘挺, 李生

(哈尔滨工业大学计算机学院信息检索研究室, 哈尔滨 150001)

E-mail: {car, tliu, lisheng}@ir.hit.edu.cn

摘要

全自动的语义分析一直是自然语言理解的主要目标之一。通过深层语义分析, 可以将自然语言转化为形式语言, 从而使计算机能够与人类无障碍的沟通。为达此目的, 人们已经进行了多年的努力, 然而由于这一问题过于复杂, 目前取得的效果并不理想。浅层语义分析是对深层语义分析的一种简化, 它只标注与句子中谓词有关的成份的语义角色, 如施事、受事、时间和地点等。其能够对问答系统、信息抽取和机器翻译等应用产生推动作用。语义角色标注是浅层语义分析的一种实现方式, 具有定义清晰, 便于评价的优点。本文描述了语义角色标注现有的语料库资源, 各种分析方法等。特别对该任务的一些难点问题提出了具体的解决方案, 并对语义角色标注的发展给出了一个初步展望。

关键词: 自然语言理解; 浅层语义分析; 语义角色标注;

1. 引言

语言分析一般分为三个层次: 句法、语义、语用。句法分析关心的是词语如何排列形成正确的句子, 并决定每个词语在句子中充当的结构角色。句法分析问题早已引起人们的广泛关注, 并取得了积极的进展。所谓语义分析, 指的是将自然语言句子转化为反映这个句子意义(即句义)的某种形式化表示。即将人类能够理解的自然语言转化为计算机能够理解的形式语言, 做到人与机器的互相沟通。而语用分析则研究影响语言行为(如招呼、劝说)的标准和支配轮流发言的规则, 目前在自然语言处理领域还鲜有研究。

对句子进行深入的语义分析, 一直是从事自然语言处理研究的学者们追求的主要目标。然而由于技术水平等于原因的限制, 语义分析一直局限于受限领域。随着通用领域句法分析等技术的进步, 近来语义分析开始触及更宽的领域。但是通用的深层语义分析仍然很难做到, 所以目前人们更关心“浅层语义分析(Shallow Semantic Parsing)”。它不考虑时态信息, 例如“他将来北京。”与“他来北京了。”, 虽然时态并不相同, 但是浅层语义表示是相同的。同时, 也不考虑谓词改变但语义不变的情况, 例如“曹雪芹写了《红楼梦》”与“《红楼梦》的作者是曹雪芹”, 虽然它们的语义相同, 但是浅层语义分析的结果并不一样, 需要根据具体的应用进行更深入的处理。最后, 浅层语义分析也不考虑人、物的指代等情况。

语义角色标注(Semantic Role Labeling)是浅层语义分析的一种实现方式, 具有分析任务定义明确, 便于评价等优点。该方法并不对整个句子进行详细的语义分析, 而只是标注句子中的一些成份为给定谓词(动词、名词等)的语义角色(论元), 这些成分作为此谓词框架的一部分被赋予一定的语义含义, 例如“[委员会 Agent][明天 Tmp]将要[通过 v][此议案 Patient]。”其中, “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。这句话也可以表示成:

[明天 Tmp] [委员会 Agent]将要[通过 v][此议案 Patient]。

[明天 _{Temp}] [此议案 _{Patient}] 将要 被 [委员会 _{Agent}] [通过 _v].
[此议案 _{Patient}] [明天 _{Temp}] 将要 被 [委员会 _{Agent}] [通过 _v].

.....

这些句子虽然形式不同，即句法分析结果不同，但是它们的含义是一样的，也就是说谓词“通过”的各个语义角色相同。

目前，语义角色标注已经被成功的应用于问答系统^{[1][2]}、信息抽取^[3]、自动文摘^[4]、文本蕴涵^[5]、词义消歧^[6]、信息检索^[7]、指代消解^[8]、机器翻译^[9]、生物信息学^[10]等领域。以问答系统为例，语义角色标注指明某一活动发生的时间、地点等语义角色，自然的可以对这些类问题进行回答。因此目前语义角色标注引起了越来越多的重视。

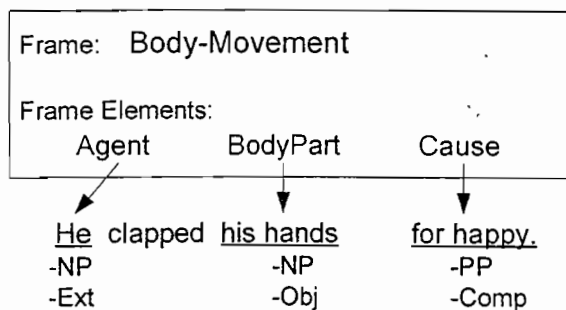
另外，国际上曾经举行过五次语义角色标注评测，依次为 Senseval-3^[11]，CoNLL (Conference on Computational Linguistics Learning) 2004^[12]、2005^[13]共享任务(Shared Task)、SemEval 2007: Task 17^[14]以及最近的 CoNLL 2008 共享任务^[15]。和其它自然语言处理领域一样，历次国际评测的成功举办，也极大的推动了语义角色标注研究的发展。这些评测不但为参赛组织和机构提供了系统性能比较和方法共享的平台，还向所有研究人员提供了标准的评测数据和工具，进一步吸引了更多的人参与到语义角色标注研究中来。

本文在接下来的部分首先介绍目前较为成熟的语义角色标注语料库以及这些语料库各自的特点；然后，我们给出了目前进行语义角色标注的一些方法；最后，展望了今后语义角色标注技术的发展趋势。

2. 语料资源

要想进行语义角色标注，需要好的语料资源的支持。目前，英语较为知名的语义角色标注资源为 FrameNet^{[16][17]}、PropBank^{[18][19]}和 NomBank^{[20][21]}。

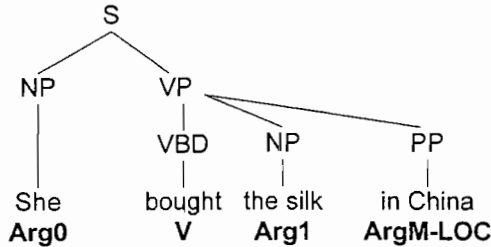
其中，U.C.Berkeley 开发的 FrameNet 以框架语义^[22]为标注的理论基础对英国国家语料库进行标注。它试图描述一个词汇单元(动词和部分名词以及形容词)的框架，同时也试图描述这些框架之间的关系。图 1 是 FrameNet 中表示身体动作的一个语义框架以及用其对一个句子的标注实例。



图(1): FrameNet 框架以及句子标注

PropBank 是 UPenn 在 Penn TreeBank 句法分析的基础上，标注的语义角色信息。与 FrameNet 不同的是，PropBank 只对动词(非系动词)进行标注，相应的被称作目标动词。而且只包含 50 多个语义角色，相同的语义角色由于目标动词不同会有不同的语义含义，其具

体解释在 PropBank 中的 Frame 文件给出。其中核心的语义角色为 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 表示时间等等。图 2 是 PropBank 中对一个句子的标注实例。



图(2): PropBank 中句子标注

与 FrameNet 相比，PropBank 基于 Penn TreeBank 手工标注的句法分析结果，因此标注的结果几乎不受句法分析错误的影响，准确率较高；而且它几乎对 Penn TreeBank 中的每个动词及其语义角色进行了标注，因此覆盖范围更广，可学习性更强。

与 PropBank 标注 Penn TreeBank 中的动词做谓词不同，NomBank 标注了其中的名词作谓词的情况，参数的类别和表示同 PropBank 是一样的。例如：名词短语“John’s replacement Ben”和“Ben’s replacement of John”中，名词 replacement 便是谓词，Ben 是 ARG0，表示替代者，John 是 ARG1 表示被替代者。

除英语外，许多其它语言也建立了各自的语义角色标注库，例如：SALSA^[23]是德语版的 FrameNet；Prague Dependency Treebank^[24]项目进行了大量的句法和语义标注(捷克语)，甚至包括指代消解的标注等；Chinese PropBank^[25]以及 Chinese NomBank^[26]分别是基于 Chinese Penn TreeBank^[27]标注的，和 PropBank 以及 NomBank 对应的汉语语义角色标注资源。Chinese FrameNet(CFN)^[28]则是山西大学参考英文 FrameNet 构建的中文框架语义词典。

3.自动语义角色标注方法

有了语料库资源作为基础，接下来的，也是最主要的任务就是从这些资源中学习已有知识，然后进行自动的语义角色标注。

3.1 标注过程

通常的语义角色标注分为 3 个步骤：谓词的识别与分类、语义角色的识别与分类以及后处理。

其中谓词的识别与分类是识别句子中的谓词并对多义的谓词赋予其适当的词义。可以采用简单的规则进行谓词的识别与分类，但通常基于统计的方法更有效。

语义角色的识别过程一般是对一个标注单元(句法成分、短语等，详细介绍见下节)是否是语义角色加以判别，并保留识别成语义角色的标注单元，待进一步分类具体属于某类语义角色，这样可以减少进入分类判别的实例个数，加快处理速度。有时在语义角色识别之前需要根据一些启发式规则进行剪枝，删除大部分不可能成为语义角色的句法成分^[29]，

这样可以大幅减少输入到分类器中的实例个数，提高训练和测试的效率。

最后根据语义角色标注本身的一些固有约束进行后处理。这些约束包括，一个谓词不能有重复的语义角色并且语义角色不存在嵌套等等。可以简单的直接利用这些规则，也有学者使用整数线性规划的算法^[30]将这些规则融入统一的框架之中，并获得了较好的效果。

3.2 标注基本单元

语义角色标注的基本单元可以是句法成分(Constituent)、短语(Phrase)、词(Word)或者依存关系(Dependency Relation)等等。

一般认为语义角色是与某一句法成分相对应的。也就是说一个语义角色必然对应着一个句法成分，反之未必。因此，现在多数的语义角色标注系统通常都是以句法成分为基本的标注单元。这种策略，在句法分析比较成熟的语言(如英文等)上表现得较好。

然而，在其它语言或者新的领域(如生物文本)上，很难自动的获得这种深层句法分析的结果，而且现有的句法分析系统，在通用领域表现欠佳。为此有人试图将语义角色标注建立在浅层句法分析的基础之上^[31]，毕竟浅层句法分析的鲁棒性要好于深层句法分析。

词是比短语更细的语言单位，有些语义角色标注系统也使用其作为标注的基本单位，然而并不如基于短语和句法成分的效果理想^[32]。

除此以外，Hacioglu 使用依存句法分析结果进行浅语义角色标注^[33]，也取得了与基于短语结构句法分析方法相近的结果。特别是 CoNLL2008 的评测，重点就是基于依存句法分析的语义角色标注。因此，基于依存句法分析的语义角色标注必将成为今后的重要研究方向。

3.3 统计学习方法

我们可以将识别和分类步骤(包括谓词和语义角色的识别和分类)看成是一个基于复杂条件的分类问题。最初解决此类问题是由人类专家根据自身的经验总结规则知识，构建知识库，再使用这些知识进行分类。然而，这种方法不但需要专业领域专家的知识 and 花费专家们大量的时间，而且规则之间会产生冲突。

为了克服知识库方法的缺点，人们后来使用统计学习的方法。该方法不需要有专业技能的专家书写知识库，只需要有一定专业知识的人对数据按照学习目标进行一定的标注；然后以此为训练数据，构造统计学习模型；再使用各种统计学习方法对模型参数加以求解。

基于特征的学习方法

目前，基于特征的方法被认为是较为有效的统计学习方法，它一般是根据一定的语言学知识提取各种特征并构成特征向量，然后使用某种学习算法对这些实例进行自动的学习，并对新的实例进行预测。

语义角色标注中经常使用的统计学习方法包括最大熵^[34]、支持向量机(SVM)^[35]、感知器^[36]等。一般的，这些方法又被称作基于特征的方法，因为在使用它们的时候，需要人为的定义大量具有区分度的特征模板，然后根据此模板将每个实例转化为特征向量进行学习或预测操作。

在语义角色标注中，有一些基本的语言学特征对系统的性能有着直接的影响。这些特征及其说明举例如下：

谓词：即待标注的目标动词，通常取其原形；

路径：指的是待标注的单元与目标动词之间句法分析树上的路径。

短语类型：待标注的单元所属短语的类型，如 NP、VP、PP 等；

位置：待标注的单元与目标动词之间的相对位置，这是二元特征，分别为“前”和“后”；

语态：目标动词的语态二元特征，分别为“主动”和“被动”；

核心词：标注单元的核心词；

动词子类框架：记录了目标动词及其子节点的结构信息。

以上这些特征，都从不同的侧面反映了待标注单元的语义角色信息。但是这些仅仅反映了待标注单元的局部信息。一些全局特征，如标注单元语义角色的序列特征，全句含有的目标动词数目等，也可以作为特征加入，具体可以采用语言模型^[35]、利用已标注语义角色特征^[37]、条件随机域^[38]等序列标注方法。

基于 Kernel 的学习方法

基于 Kernel 的机器学习方法，是对上面基于特征的学习方法的一种扩展。其基本思想是将低维线性不可分问题映射到高维空间，变为线性可分问题。通常这种映射可以通过计算 Kernel 函数隐式的达到，从而降低时间和空间复杂性。Kernel 函数的计算能很好的融入支持向量机、感知器等学习算法，使它们可以较好的处理线性不可分问题。

在自然语言处理领域，研究者们又将 Kernel 的思想升华，使其能够更好的处理结构化的数据，发掘更多的结构化特征。最早由 Haussler^[39]以及 Watkins^[40]提出使用动态规划的方法高效计算结构化数据的 Kernel 函数，这又被称作 Convolution Kernel。在此之后，越来越多的学者利用相似的思想于自己的研究工作，如句法分析^[41]，文本分类^[42]等等。特别是 Moschitti 通过计算句法分析树的 Kernel 函数进行语义角色标注^[43]。Che 等进一步改进 Moschitti 的方法，并提出了一种新的 Kernel 计算方法—混合卷积树 Kernel^[44]，该方法更适合语义角色标注问题。通常的树 Kernel 函数计算要求子树必须完全相同，然而这一要求有时过为苛刻，例如在两个句法分析树中的“NP->DT JJ NN”(NP->a red car)和“NP->DT NN”(NP->a car)对标准树 Kernel 函数计算没有贡献，尽管他们有很相似的句法结构属性，并且在一个给定的谓词中应该扮演相似的语义角色。因此，Zhang 等设计了句法驱动的树 Kernel 计算方法^[45]，进一步提高了语义角色标注系统的性能。

3.4 系统融合方法

语义角色标注在很大程度上依赖句法分析系统的性能，然而目前单一句法分析系统的性能并非理想，因此如何融合更多句法分析的结果，成为提高语义角色标注系统性能的一个非常重要的方法。在 CoNLL2005 评测中，排在前面的参赛队伍，均采用了这一思想。

已有的具体方法包括：1)基于规则的方法^[46]；2)基于整数线性规划的方法^[47]；3)基于重排序的方法^[48]；4)基于机器学习的方法^[49]；5)基于组块分析的方法^[50]等等；

以上的各种方法，都或多或少的提高了系统的整体性能，当然也造成了系统构造复杂，分析效率不高等问题，还有待进一步解决。

4. 研究展望

语义角色标注是对语义分析的一个有益的探索，为了进一步提高其性能，并使其对实

际应用产生更大的帮助，我们认为目前有一下的工作亟待解决：

4.1 特征工程

多年的研究结果表明，影响机器学习系统性能的首要因素并非机器学习模型，而是使用的特征。近来语义角色标注的研究也揭示了这一点。因此，若想提高系统的性能，细致的特征工程工作是必不可少的。

然而特征向量的形式往往很难恰当的表达结构化的特征，基于树 Kernel 的学习方法是解决该问题的一种有益尝试，然而该研究方向才刚刚开始，还有很多问题有待进一步研究。

词汇化(包括谓词以及语义角色)特征对于语义角色标注也尤为重要，但是该特征过于稀疏，需要进行适当的聚类。例如，“我爱吃苹果。”与“我爱吃香蕉。”中“苹果”和“香蕉”都是“吃”的受事，如果能将类似的词语进行聚类，就可以提高系统的泛化能力，使用词义等消歧技术也能达到类似的功能。另外，Kernel 函数是一种特殊的相似度函数，如何将词汇语义的相似度作为 Kernel 函数，利用基于词汇语义 Kernel 的方法进行语义角色标注也是今后的一个研究方向。

4.2 领域自适应

为了更好的应用语义角色标注，必须解决领域自适应问题，也就是说解决测试语料和训练语料属于不同的领域，其性能下降较多的问题。这种现象在历次 CoNLL 评测中的表现尤为明显，Brown 语料上的评测结果均较之 WSJ 语料结果低 10%左右。但是，我们不可能针对不同的领域，都标注大量的训练语料，所以领域的自适应成为一种必需。具体的方法可以参考机器学习领域的一些最新研究进展，如半指导学习^[51]，迁移学习^[52]等。

4.3 句法和语义分析互动

目前人们普遍认为语义角色标注在很大程度上依赖句法分析的结果，为了尽量减小由于句法分析错误对语义角色标注造成的不利影响，人们曾经试图融合多个句法分析的结果，并取得了积极的进展。但是这种提高是单方向的，也就是说仅仅提高了语义角色标注的性能，可否利用语义角色标注帮助句法分析目前还鲜有人研究。通过在此方向上的探索可以使得句法和语义分析充分互动，互为补充。

4.4 语义依存分析

目前，语义角色标注结果的表示形式还存在一些不足：

- 1、 语义角色的名称和种类不统一，命名过于随意。在 PropBank 中，虽然使用了统一的 ArgX 形式命名语义角色，但是相同的语义角色对于不同的谓词，其具体的含义(见谓词的 Frame 文件)是不同的，对于这些含义的解释往往采用自然语言的描述形式，非常随意，给更深层的应用带来一些困难。例如在中文 PropBank 中，不同的描述竟有 3,000 多种，其中 2,500 多种描述仅出现了一次。
- 2、 语义角色作为统一的整体，没有考虑其内部的结构。目前的语义角色标注仅仅考虑了谓词及其语义角色之间的关系，而忽略了语义角色内部的结构关系。这也是为什么语义角色标注被归为浅层语义分析的一个主要原因。对于某些应用，往往

需要考虑这些更细致的关系。

综上，我们认为语义依存分析是一种更合适的语义分析表示方式。它将依存句法分析和语义分析相结合，将依存的句法关系转化为语义关系，不同于 PropBank 中对语义角色的描述，我们可以提取数量不多的抽象语义关系。

另外，我们认为这种分析对中文会尤为有效，因为其可以充分利用中文意合且缺乏形态变化的特点，跨越较难的中文句法依存分析步骤，直接进行语义依存分析，而且分析结果不是浅层的语义分析，而是分析出一棵完整的语义依存树，将会更加有助于后续的应用研究。

参考文献

- [1] S. Narayanan, S. Harabagiu. Question Answering Based on Semantic Structures. Proceedings of Coling 2004. 2004
- [2] D. Shen, M. Lapata. Using Semantic Roles to Improve Question Answering. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). 2007:12–21.
- [3] M. Surdeanu, S. Harabagiu, J. Williams, et al. Using Predicate-argument Structures for Information Extraction. Proceedings of ACL 2003. 2003
- [4] Y. W. Y. L. A. S. Gabor Melli, Zhongmin Shi, F. Popowich. Description of Squash, the Sfu Question Answering Summary Handler for the DUC-2006 Summarization Task. Proceedings of the Document Understanding Conference 2006 (DUC-2006). 2006
- [5] Rodrigo de Salvo Braz, Roxana Girju, Vasin Punyakanok, Dan Roth, and Mark Sammons. An inference model for semantic entailment in natural language. In Proceedings of AAAI 2005.
- [6] Dang, H. T. and Palmer, M. 2005. The role of semantic roles in disambiguating verb senses. In Proceedings of ACL-2005. 42-49.
- [7] M. W. Bilotti, P. Ogilvie, J. Callan, et al. Structured Retrieval for Question Answering. SIRIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA, 2007:351–358
- [8] Simone Paolo Ponzetto, Michael Strube: Semantic Role Labeling for Coreference Resolution. EACL 2006
- [9] J. Hajic, M. Cmejrek, B. Dorr, et al. Natural Language Generation in the Context of Machine Translation. Tech. rep., Center for Language and Speech Processing, Johns Hopkins University, Baltimore, 2002
- [10] Bethard S, Lu Z, Martin JH, Hunter L: Semantic Role Labeling for Protein Transport Predicates, BMC Bioinformatics, 9:277, 2008
- [11] <http://www.cs.unt.edu/~rada/senseval/senseval3/workshop.html>
- [12] X. Carreras and L. Màrques. Introduction to the conll-2004 shared task: Semantic role labeling. In Proceedings of CoNLL-2004, pages 89-97, Boston, MA, USA, 2004.
- [13] <http://www.lsi.upc.edu/~srlconll/>

- [14] <http://nlp.cs.swarthmore.edu/semEval/>
- [15] <http://www.yr-bcn.es/conll2008/>
- [16] <http://framenet.icsi.berkeley.edu/>
- [17] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet project. In C. Boitet and P. Whitelock, editors, Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics, pages 86-90, San Francisco, California, 1998. Morgan Kaufmann Publishers.
- [18] <http://www.cis.upenn.edu/~ace/>
- [19] M. Palmer, D. Gildea, and P. Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles, Computational Linguistics, 31(1), 2005.
- [20] <http://nlp.cs.nyu.edu/meyers/NomBank.html>
- [21] Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. Annotating noun argument structure for NomBank. In Proceedings of LREC-2004, pages 803-806, Lisbon, Portugal.
- [22] J. Fillmore. Frames and the semantics of understanding. Quaderni di Semantica, IV(2), 1985.
- [23] K. Erk, A. Kowalski, S. Pado, and M. Pinkal. Towards a resource for lexical semantics: A large german corpus with extensive semantic annotation. In Proceedings of at ACL 2003, Sapporo, 2003.
- [24] E. Hajicova. Prague dependency treebank: From analytic to tectogrammatical annotation. In Proceedings of the First Workshop on Text, Speech, Dialogue, pages 45-50, 1998.
- [25] <http://www.cis.upenn.edu/~chinese/cpb/>
- [26] N. Xue. 2006. Annotating the predicate-argument structure of Chinese nominalizations. In Proceedings of the fifth international conference on Language Resources and Evaluation, Genoa, Italy.
- [27] <http://www.cis.upenn.edu/~chinese/>
- [28] Liping You, Kaiying Liu. Building Chinese FrameNet Database. In Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering, 2005
- [29] N. Xue, M. Palmer. Calibrating Features for Semantic Role Labeling. Proc. EMNLP 2004, 2004
- [30] V. Punyakanok, D. Roth, W. Yih, et al. Semantic Role Labeling via Integer Linear Programming Inference. COLING-2004, Geneva, Switzerland, 2004, 1346-1352.
- [31] K. Hacioglu, S. Pradhan, W. Ward, et al. Semantic Role Labeling by Tagging Syntactic Chunks. HLT-NAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004), 2004, 110-113
- [32] Y. Kwong, B. K. Tsou. Semantic Role Tagging for Chinese at the Lexical Level. Proceedings of IJCNLP 2005, 2005
- [33] K. Hacioglu. Semantic Role Labeling Using Dependency Trees. Proc. Coling 2004, 2004

- [34] Ting Liu, Wanxiang Che, Sheng Li, Yuxuan Hu, and Huaijun Liu. Semantic role labeling system using maximum entropy classifier. In Proceedings of CoNLL-2005, 2005.6, pp189-192
- [35] S. Pradhan, K. Hacioglu, V. Krugler, et al. Support Vector Learning for Semantic Argument Classification. Machine Learning Journal. 2005
- [36] X. Carreras, L. Marquez, G. Chrupa?a. Hierarchical Recognition of Propositional Arguments with Perceptrons. H. T. Ng, E. Riloff, (Editors) HLTNAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004), 2004, 106–109
- [37] Z. P. Jiang, J. Li, H. T. Ng. Semantic Argument Classification Exploiting Argument Interdependence. Proceedings of IJCAI-2005. 2005
- [38] Kristina Toutanova. Effective statistical models for syntactic and semantic disambiguation. PhD Thesis, Stanford CS Department, 2005.
- [39] D. Haussler. Convolution Kernels on Discrete Structures. Tech. Rep. UCSC-CRL-99-10, 1999.
- [40] C. Watkins. Dynamic Alignment Kernels. Tech. Rep. CSD-TR-98-11, 1999.
- [41] M. Collins, N. Duffy. Convolution Kernels for Natural Language. Proceedings of NIPS-2001, 2001.
- [42] H. Lodhi, C. Saunders, J. Shawe-Taylor, et al. Text classification using string Kernels. J. Mach. Learn. Res. 2002, 2:419–444
- [43] Moschitti. A Study on Convolution Kernels for Shallow Semantic Parsing. Proc. ACL 2004, 2004
- [44] Wanxiang Che, Min Zhang, Ting Liu, Sheng Li. A Hybrid Convolution Tree Kernel for Semantic Role Labeling. In Proceedings of ACL2006. Sydney. July 2006
- [45] Min ZHANG,Wanxiang CHE,Ai Ti AW,Chew Lim TAN,Guodong ZHOU,Ting LIU,Sheng LI, A Grammar-driven Convolution Tree Kernel for Semantic Role Classification,ACL2007
- [46] Lluís Màrquez, Pere R. Comas, Jesús Giménez and Neus Català, Semantic Role Labeling as Sequential Tagging. CoNLL-2005
- [47] Vasin Punyakanok, Peter Koomen, Dan Roth and Wen-tau Yih, Generalized Inference with Multiple Semantic Role Labeling Systems.
- [48] Aria Haghighi, Kristina Toutanova and Christopher Manning, A Joint Model for Semantic Role Labeling.
- [49] Surdeanu M., Màrquez L., Carreras X., and Comas P. Combination Strategies for Semantic Role Labeling. Journal of Artificial Intelligence Research, 29, 2007.
- [50] Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James H. Martin and Daniel Jurafsky, Semantic Role Chunking Combining Complementary Syntactic Views.
- [51] Chapelle, O., B. Sch?lkopf and A. Zien: Semi-Supervised Learning. MIT Press, Cambridge, MA (2006).
- [52] Do, Cuong B. and Andrew Y. Ng. Transfer learning for text classification. NIPS-2005