

基于机器翻译的跨语言关系抽取*

胡亚楠, 舒佳根, 钱龙华, 朱巧明

(苏州大学自然语言处理实验室, 江苏, 苏州, 215006)

(苏州大学计算机科学与技术学院, 江苏, 苏州, 215006)

{20114227025, 20124227015, qianlonghua, qmzhu}@suda.edu.cn

摘要: 训练语料库的规模对基于机器学习的命名实体间语义关系抽取具有重要的作用, 而语料库的人工标注需要花费大量的时间和人力。本文提出了使用机器翻译的方法将源语言的关系实例转换成目标语言的关系实例, 并通过实体对齐策略将它们加入到目标语言的训练集中, 从而使资源丰富的源语言帮助欠资源的目标语言进行语义关系抽取。在 ACE2005 中英文语料库上的关系抽取实验表明, 无论是将中文翻译成英文, 还是将英文翻译成中文, 都对另一种语言的关系抽取具有帮助作用。特别是当目标语言的训练语料库规模较小时, 这种帮助就尤其显著。

关键词: 跨语言关系抽取; 机器翻译; 实体对齐

Cross-Lingual Relation Extraction via Machine Translation

Hu Ya'nan, Jiagen Shu, Qian Longhua, Qiaoming Zhu

(Natural Language Processing Lab, Soochow University, Suzhu, Jiangsu, 215006)

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

{20114227025, 20124227015, qianlonghua, qmzhu}@suda.edu.cn

Abstract: The scale of training corpus plays an important role in machine learning-based semantic relation extraction between named entities, however, the annotation of corpus is time-consuming and labor-intensive. In order that a resource-rich language can help a resource-poor language in semantic relation extraction, we propose an approach to transforming relation instances from the source language to the target language via machine translation, and then add them into the training corpus of the target language by way of entity alignment. The experiments on the ACE2005 Chinese and English corpora show that, Chinese and English can help each other in relation extraction. Furthermore, this help is particularly significant especially when the scale of training corpus in target language is small.

Keywords: Cross-lingual Relation Extraction; Machine Translation; Entity Alignment

1 引言

命名实体间语义关系抽取(简称实体关系抽取, 或关系抽取)是自然语言处理(NLP)和信息抽取中的一个重要环节, 其任务是从自然语言文本中提取出两个命名实体之间所存在的语义关系。例如短语“我国 国际 机场”中的两个实体“我国”(GPE.Special)和“机场”(FAC.Airport)之间存在的部分整体关系(PART-WHOLE.Geographical)。关系抽取对问题回答、文本摘要、信息融合和知识挖掘等自然语言处理应用都具有重要的意义。

主流的关系抽取研究都采用机器学习的方法, 根据其对标注语料库规模的不同需求, 分为指导性学习^[1-3]、弱指导学习^[4]和无指导学习^[5]等方法。标注语料的数量和质量通常决定了抽取性能的好坏, 但语料的标注费时费力, 因此近几年涌现出一种新的学习方法--自指导学习方法^[6], 其特点在于训练语料不是由人工标注, 而是通过将知识库中存在的关系实例自动映射到自然语言相应文本的方法来产生。不过, 虽然它解决了语料的人工标注问题, 但由于产生的关系实例中存在着噪音, 因此抽取性能尚不够理想。

*基金资助: 国家自然科学基金(60873150, 90920004)、江苏省自然科学基金(BK2010219)、江苏省高校自然科学基金重大项目(11KJA520003)。

另一方面，目前存在着用多种语言标注的关系抽取语料库，如 ACE 2005 中的中文和英文语料库，而不同的语言之间具有一定的相似性与互补性，因此是否可以用一种语言的标注语料库来帮助另一种语言的关系抽取呢？特别是当一种语言的标注语料资源较丰富，而另一种语言的标注语料资源较贫乏时，这种帮助作用就显得更为重要。

本文的思想是使用机器翻译的方法，将关系实例从源语言翻译为目标语言，用于帮助目标语言进行关系抽取。由于跨语言关系抽取不仅涉及到实体和实例的翻译，还涉及到源语言和目标语言之间的实体对齐，因而与其它跨语言 NLP 任务^[7]相比，更为复杂和困难。在 ACE2005 中文和英文语料库上的实验表明，尽管机器翻译过程中会带来一定的噪音，但我们的方法仍能提高目标语言（中文或英文）的关系抽取性能。特别是当目标语言的训练语料较少时，关系抽取性能的提高就尤为显著。

本文的第 2 节回顾了相关工作。第 3 节介绍了本文使用的中英文关系抽取的特征集。第 4 节提出了基于机器翻译的实例映射方法。第 5 节说明本文实验所使用的语料及实验设置，并给出实验结果和分析。最后第 6 节为总结和展望。

2 相关工作

在关系抽取研究方面，大量的研究都集中于单语言的关系抽取，英文方面典型的方法有基于特征的方法^[8,1]和基于核函数的方法^[9-10]，中文方面也有这两种方法的具体应用^[11-12]。

而在跨语言的关系抽取方面，相关研究工作还非常少。Kim 等（2010）^[7]提出了一种基于标注映射的跨语言关系抽取方法，它将源语言训练得到的模型应用到平行语料库的源语言部分，那些识别出来的可靠性较高的源语言实例再映射到它们的目标语言部分，从而得到目标语言的训练实例，最后进行目标语言的关系检测。Kim 和 Lee（2012）^[13]则进一步使用了一种基于图的半监督学习算法，即标签传播算法，通过迭代方法将源语言中更多的上下文信息映射到目标语言中，从而提高从源语言映射到目标语言的关系实例数量和质量。与他们的研究工作不同，我们使用机器翻译方法对源语言中已标注的关系实例进行直接翻译，而不是使用平行语料库进行实例映射。并且，我们对 ACE 语料库进行了通用关系的抽取，而不仅仅是关系检测或少量关系类型的抽取。

除了机器翻译本身，在自然语言处理的其它领域中，跨语言处理也获得了广泛的应用。早期主要用于词性标注^[14]、命名实体识别^[15]和动词分类^[16]等简单任务，近期也逐渐被应用到诸如依存分析^[17]和语义角色标注^[18]等复杂任务中。其中，与我们工作有相似之处的是 Oh 等（2009）^[4]提出的基于双语协同训练的名词上下位关系分类。该方法通过翻译一种语言中可信度较高的名词语义关系实例来扩充另一种语言的训练集，并重复两种语言之间的翻译和训练过程，以期同时提高两种语言的分类性能。与我们采用机器翻译系统进行关系实例翻译不同，他们只是通过简单的词典查找来实现词汇翻译。

3 中英文关系抽取特征集

本文使用基于特征向量的机器学习方法来实现中英文关系抽取。由于本文着重于跨语言关系抽取，因此只使用了相对简单的词汇和实体信息等特征，而没有使用复杂的句法特征。这样做的另一个原因是翻译文本的语法较差，不适合用于句法分析。本文使用特征如下：

a) 实体词汇特征及上下文信息，包括实体本身和两个实体之间的词汇。我们没有使用实体左边和实体右边的词汇，这是由于初步实验结果表明，添加这些信息并不能提高关系抽取的性能，同时也不利于后期的机器翻译。词汇特征包括：

WM1: 实体 1 的词包

HM1: 实体 1 的中心词

WM2: 实体 2 的词包

HM2: 实体 2 的中心词

HM12: 实体 1 中心词和实体 2 中心词的组合

WBNUL: 实体对之间不存在任何单词

WBFL: 实体对之间仅存在一个单词

WBF: 当实体对之间至少有两个单词, 实体对之间的第一个单词

WBL: 当实体对之间至少有两个单词, 实体对之间的最后一个单词

WBO: 当实体对之间至少有三个单词, 实体对之间除第一个和最后一个之外的其它单词

b) 实体类型及其组合特征, 包括:

ET12: 两个实体的实体大类的组合

EST12: 两个实体的实体小类的组合

EC12: 两个实体的实体参照类别的组合

c) 实体参照方式, 征包括:

ML12: 两实体参照方式的组合

MT12: 两实体 LDC 参照类型的组合

d) 交叠特征, 主要用于反映实体之间的位置关系, 属于结构化信息, 包括:

#WB: 两个实体间词语的数目

#MB: 两个实体间包含其他实体参照的数目

4 基于机器翻译的实例映射

为了用一种语言(称为源语言)的标注语料帮助另一种语言(称为目标语言)的关系抽取, 本文采用基于机器翻译的实例映射方法: 首先将源语言中的关系实例通过机器翻译映射为目标语言的关系实例, 然后将这些关系实例添加到目标语言的训练集中, 最后采用统计机器学习方法进行目标语言的关系抽取。其中, 实例映射主要包含以下三个过程:

◆ 基于五倍交叉的实例过滤

从理论上讲, 在源语言中标注的关系实例都是正确和可靠的, 但实际上在统计机器学习方法中, 那些表达较复杂且稀疏的关系实例通常是不能被正确识别的, 并且由于机器翻译也存在错误, 因此由这些源语言的实例翻译所得到的目标语言实例也往往是不可靠的, 所以在实例翻译之前我们先过滤掉这些不可靠的源语言关系实例。

一种简单的过滤方法是依据关系实例的语言表达形式的复杂程度, 如两个实体之间的词汇距离、在句法树中的距离及依存路径中的距离等, 但这些度量值无法准确反映实例的可靠性。我们采用基于五倍交叉的可靠性度量方法, 其原理类似于测试集的五倍交叉验证法, 基本思想是“*当一个关系实例能借助自身语言的其它关系实例正确识别时, 它的可靠性较高, 反之则可靠性较低*”。具体方法是: 首先, 源语言的所有标注语料被平均分成五份, 其中四份用于分类模型的训练, 然后对另外一份进行测试。这样每一个关系实例都有一个预测结果, 当预测结果正确时, 说明该实例可靠性较高, 因而它被保留下来, 反之则过滤掉。另一方面, 由于关系抽取中普遍存在着数据不平衡的现象, 即负例数量远大于正例数量, 因此源语言中的负例也被过滤掉。

◆ 基于 Google Translator 的实体和实例翻译

实体和实例翻译是将源语言的实体参照和关系实例翻译成目标语言的实体参照和关系实例。我们不对源语言文本中的整个句子进行翻译, 而只针对两个实体参照之间的短语部分进行翻译。虽然这样做丢失了关系实例的上下文信息, 但短语翻译在一定程度上可提高翻译质量, 且大部分的关系实例与上下文信息并不相关, 初步的实验结果也表明了这一点。我们

采用 Google Translator¹进行实体和实例的翻译，其翻译结果也是可接受的。在翻译完之后，为了便于实体对齐，需要对英文译文进行符号化处理，对中文译文则采用 ICTCLAS²进行分词处理。

◆ 基于混合匹配原则的实体对齐

实体对齐的目的是将源语言实体参照匹配到其译文的实体参照，即在目标语言的关系实例中自动标注出相应的实体参照，从而实现目标语言关系实例的特征向量构造。实体对齐是跨语言关系抽取中的关键问题，其难点在于实体参照在不同的上下文中可能被翻译成不同的译文，即单独翻译的实体参照和在关系实例中翻译的实体参照有可能不同，这给实体对齐带来了很大的难度。比如实体参照“官员”被单独翻译成“Officer”，而在关系实例“叙利亚 官员”中则被翻译成为“officials”。

针对这一情况，我们提出了基于混合原则的实体对齐方法。首先是完全匹配原则，如果目标语言的实体参照 (M_d) 能够在目标语言的关系实例 (R_d) 中找到相同内容，则实体的完全匹配成功；其二是部分匹配原则，如果从 M_d 的右部开始能够在 R_d 中找到部分相同内容，则实体的部分匹配成功；最后是基于翻译词表的对齐原则，如果上述两种方法都不能实现实体匹配，则尝试借助于翻译词表直接从源语言的实体参照 (M_s) 到目标语言的关系实例 (R_d) 进行匹配。翻译词表 L 从中英文平行语料库 FBIS 中导出，每项为 (s_i, d_i, p_i) ，分别表示源语言词汇、目标语言词汇和翻译概率。匹配过程为：对于 M_s 中的任一单词 s_i 和 R_d 的任意一个单词 d_j ，如果能在词表 L 中找到表项 (s_i, d_j, p_j) 且翻译概率大于阈值 θ (设为 0.002)，则认为这是一个单词匹配。如果能够在 R_d 中找到 M_s 的一个连续匹配，则认为 M_d 匹配成功。

5 实验

本节首先介绍实验语料和实验设置，包括所使用的语料库、分类器及性能评估指标，然后给出实验结果和分析。

5.1 实验语料和设置

本文采用 ACE2005 中英文语料库作为语义关系抽取的实验数据。该语料库定义了实体之间的 6 个关系大类，18 个关系小类。为使得中英文实例数量在同一个级别，我们对中文数据进行了部分的删减。最后，我们使用中文语料中的 457 个文件，来源为广播新闻和新闻专线，包含 6,268 个关系正例及 69,047 个关系负例，其中有 3,069 个关系正例被翻译并对齐到英文；使用英文语料中的 498 个文件，主要来源为广播谈话、广播新闻、电话谈话等，包含 6,253 个关系正例及 72,883 个关系负例，其中有 2,679 个关系正例被翻译并对齐到中文。

语料的预处理包括从 XML 格式的 ACE 数据文件中提取出原始文本以及实体和关系的标注信息，并对文本进行分段、分句和符号化（中文则为分词）。然后，对一个句子中出现的实体参照进行两两配对以产生所有可能的关系实例。最后，提取所有关系实例的特征构造特征向量。

本文选用 SVM-Light 作为分类器，采用五倍交叉验证策略计算抽取性能的平均值。评估标准采用常用的准确率 P 、召回率 R 和 $F1$ 指数，其中 $F1$ 指数即为准确率 P 和召回率 R 的调和平均。

5.2 实验结果和分析

(1) 训练语料规模对跨语言关系抽取性能的影响

图 1 和图 2 分别给出了在不同规模的目标语言训练语料上添加翻译语料前后，中英文各自的大类关系抽取性能。其中，图 1 的三个子图 (a)、(b) 和 (c) 分别给出了在英文训练

¹ <http://translator.google.com>

² <http://ictclas.org/>

语料上添加翻译语料前后的准确率 P、召回率 R 和 F1 值，图 2 为中文语料上的实验性能。横坐标表示目标语言语料库的规模，如“1/20”表示训练实例数为总数的 1/20，纵坐标表示相应的性能。

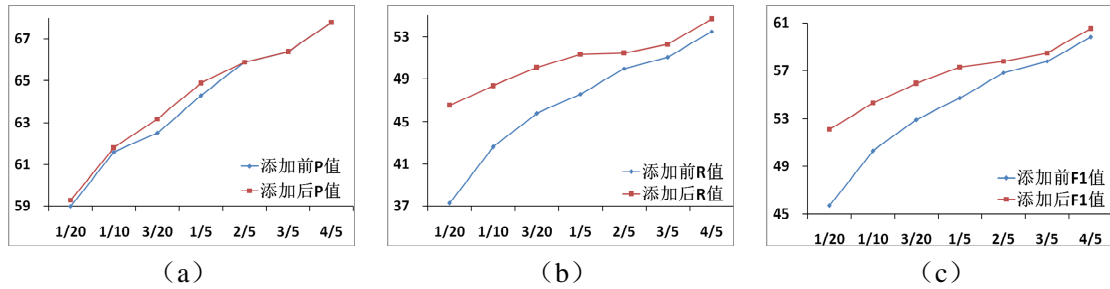


图 1 不同规模的英文语料添加翻译实例前后性能对比

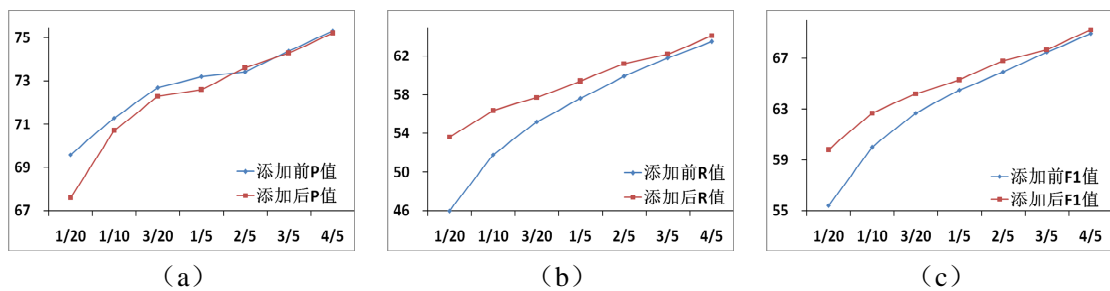


图 2 不同规模的中文语料添加翻译实例前后性能对比

从图 1 和图 2 可以看出：

- ◆ 无论是英文还是中文关系抽取，添加翻译语料后，总体性能都随目标语言训练语料规模的不同而获得了不同程度的提升，且目标语言训练语料规模越小，总体性能提高幅度越大。这说明增加从另一种语言翻译过来的关系实例后，确实有助于目标语言的关系抽取，而且当目标语言的训练语料越小时（欠资源），数据就越稀疏，这种帮助作用也就越大；
- ◆ 总体性能 F1 的提高主要来源于召回率的提高，而准确率提高则很小，甚至降低。这说明翻译得到的关系实例，除了能弥补目标语言中训练实例的不足之外，也带来了一定的噪音；
- ◆ 中英文的表现稍有不同，翻译实例对英文关系抽取的作用大于对中文关系抽取的作用，且前者的准确率能维持不变或略有提高，而后者的准确率则基本上降低。主要原因有三点，首先，由实验设置可知中文翻译并对齐到英文的实例数量多于英文翻译并对齐到中文的实例数量；其次，通常认为中文到英文的翻译质量要高于英文到中文的翻译质量，最后，翻译之后对齐之前需要对中文进行分词处理，对英文进行符号化处理，而众所周知，中文分词难度大于英文符号化。

(2) 不同类型上的跨语言关系抽取性能比较

表 1 和表 2 分别给出了在五分之一规模的目标语言语料中添加了翻译语料后，英文和中文在各个不同关系类型上的抽取性能及变化。两表的第二列分别表示英文和中文在各个关系大类上的实例数目。表 1 中的“1/5e”和“1/5e+ACE2005c”两列分别表示采用五分之一规模的英文语料和在此基础上加入从 ACE2005 中文语料库中翻译过来的实例，“Delta”表示两者之间的性能变化，其中各列性能的最大值用粗体表示。表 2 的各列与表 1 基本相同，是在中文语料上的实验结果。

表 1 1/5 英文训练语料添加翻译实例前后不同关系类型的性能对比

关系类型	#	1/5e			1/5e+ACE2005c			Delta		
		P	R	F1	P	R	F1	P	R	F1
PHYS	1,443	49.8	23.2	31.7	51.9	26.5	35.0	2.1	3.3	3.4
PART-WHOLE	938	68.3	56.3	61.7	67.2	65.5	66.3	-1.1	9.2	4.6
PER-SOC	710	77.9	59.6	67.3	79.4	60.3	68.4	1.5	0.7	1.1
ORG-AFF	1,854	69.9	63.4	66.4	71.9	64.3	67.8	2.0	0.9	1.4
ART	607	63.7	39.2	48.5	61.9	41.0	49.3	-1.8	1.8	0.8
GEN-AFF	701	58.4	41.2	48.3	59.5	48.5	53.4	1.0	7.3	5.1
平均值	6,253	64.0	47.8	54.7	65.1	51.3	57.3	1.0	3.5	2.6

表 2 1/5 中文训练语料添加翻译实例前后不同关系类型的性能对比

关系类型	#	1/5c			1/5c+ACE2005e			Delta		
		P	R	F1	P	R	F1	P	R	F1
PHYS	1,145	57.5	28.6	38.2	57.1	29.5	38.8	-0.4	0.9	0.7
PART-WHOLE	1,588	77.5	68.1	72.5	75.2	70.5	72.7	-2.3	2.4	0.2
PER-SOC	327	73.9	41.6	52.5	74.6	41.9	52.9	0.7	0.3	0.5
ORG-AFF	1,477	80.9	71.7	76.0	81.1	72.8	76.7	0.2	1.1	0.7
ART	432	60.0	32.9	42.2	64.4	41.0	49.9	4.5	8.1	7.7
GEN-AFF	1,299	77.1	66.7	71.5	75.5	67.5	71.3	-1.6	0.8	-0.2
平均值	6,268	73.2	57.6	64.5	72.6	59.4	65.3	-0.6	1.8	0.8

从表 1 和表 2 中可以看出,添加翻译语料后大部分关系类型的抽取性能获得了不同程度的提高,且这种提高主要来源于召回率,而精度提高不大,甚至有所降低,不过,中英文之间的表现却不尽相同。具体表现为:

- ◆ 从表 1 可以看出,添加了翻译语料后,英文的通用附属关系 (GEN-AFF)、部分整体关系 (PART-WHOLE) 和物理位置关系 (PHYS) 的 F1 值提高幅度最大,分别达到了 5.1、4.6 和 3.4,而另三种关系类型则提高幅度较小。
- ◆ 从表 2 可以看出,添加了翻译语料后,中文的施事关系 (ART) 在准确率、召回率、F1 值上面都有显著提升,而在其它大类上提高有限,在通用附属关系 (GEN-AFF) 大类上甚至稍有降低。
- ◆ 结合表 1 和表 2 可以看出,添加翻译语料后英文抽取性能的提升幅度普遍高于中文抽取性能的提升幅度,这是由于中文到英文的翻译质量一般要优于英文到中文的翻译质量,且中文的分词也会带来一定的噪音。

(3) 与其它中英文关系抽取系统的比较

表 3 列出了其它几个关系抽取系统在 ACE 中英文语料库上的大类抽取性能。由于所采用的语料库规模不同,且测试和验证方法也有所不同,因此实验结果仅供参考。不过,还是可以从表中看出,虽然采用特征较为简单,在相同语料下本文系统的性能与其它系统不相上下。另一个情况是 ACE2005 英文的抽取性能较 ACE2004 英文低得多,其主要原因是前者的语料库包含电话谈话文本,文体较为自由,且所定义的关系类型的一致性没有后者好。最后是 ACE2005 英文的抽取性能较 ACE2005 中文的抽取性能低很多,这与一般认为英文抽取性能优于中文抽取性能的观点不符。通过对语料数据的观察我们发现,中文关系实例的结构较为简单,具有前修饰结构的中文关系实例占总数的 60%左右,而在英文关系实例中,具有前

修饰结构的实例不足 30%。

表 3 与其它关系抽取系统的性能比较

系统	抽取方法	语料库	P (%)	R (%)	F1 (%)
本系统	特征向量	ACE2005 中文	75.2	64.1	69.2
Li 等 (2008) ^[19]	特征向量	ACE2005 中文	81.7	61.7	70.3
车万翔 (2005) ^[20]	特征向量	ACE2004 中文	76.1	70.2	73.3
Liu 等 (2012) ^[3]	卷积树核	ACE2005 中文	81.9	56.8	67.0
虞欢欢等 (2010) ^[21]	卷积树核	ACE2005 中文	75.3	60.4	67.0
黄瑞红等 (2008) ^[22]	卷积树核	ACE2007 中文	40.1	33.0	35.0
本系统	特征向量	ACE2005 英文	67.8	54.7	60.5
Zhang 等 (2006) ^[23]	复合核	ACE2004 英文	76.1	68.4	72.1
Qian 等 (2008) ^[9]	复合核	ACE2004 英文	83.0	72.0	77.1
Zhou 等 (2010) ^[24]	复合核	ACE2004 英文	82.7	73.5	77.8

6 总结和展望

本文使用基于机器翻译的方法实现跨语言关系抽取，探讨了在欠资源目标语言的情况下，添加从源语言翻译过来的语料对目标语言关系抽取结果的影响。通过实验发现，添加翻译语料能在一定程度上提高中英文关系抽取的性能，并且，目标语言的语料越少，翻译语料所具有的帮助作用就越明显。值得注意的是，添加目标语言语料，准确率和召回率都有所提升，而添加翻译语料，总体性能的提升主要得益于召回率的提升，准确率的增长并不是很明显，在中文上甚至有所降低。这说明翻译语料可以补充目标语言语料的不足，但也有可能带来一定的噪音。

下一步的研究工作将从以下两个方面展开：一是考虑使用更多的特征来捕获对关系抽取有用的信息；二是利用双语平行语料库来实现实例的映射以减少由机器翻译带来的噪音。

参考文献

- [1] GuoDong Zhou and Min Zhang. Extracting relation information from text documents by exploring various types of knowledge[J]. Information Processing and Management, 2007, 43:969-982
- [2] 奚斌, 钱龙华, 周国栋, 朱巧明, 钱培德. 语言学组合特征在语义关系抽取中的应用[J]. 中文信息学报, 2008, 22 (3) :44-63
- [3] Dandan Liu, Zhiwei Zhao, Yanan Hu, and Longhua Qian. Incorporating Lexical Semantic Similarity to Tree Kernel-Based Chinese Relation Extraction[C]//Computer School Center for Study of Language & Information. The 13th Chinese Lexical Semantics Workshop. Wuhan: Wuhan University, 2012,133-139
- [4] Jong-Hoon Oh, Kiyotaka Uchimoto, and Kentaro Torisawa. Bilingual Co-Training for Monolingual Hyponymy-Relation Acquisition[C]//ACL, Singapore, 2009:432-440
- [5] Min Zhang, Jian Su, Danmei Wang, Goudong Zhou, and Chew Lim Tan. Discovering Relations from a Large Raw Corpus Using Tree Similarity-based Clustering[C]//In Proceedings of IJCNLP, 2005:378-389.
- [6] Fei Wu and Daniel S. Weld. Open information extraction using wikipedia[C]// ACL. 2010:118-127.
- [7] Seokhwan Kim, Minwoo Jeong, Jonghoon Lee, and Gary Geunbae Lee. A Cross-lingual Annotation Projection Approach for Relation Detection[C]//Coling, Beijing, 2010:564-571
- [8] Nanda Kambhatla. Combining lexical, syntactic and semantic features with Maximum Entropy models for extracting relations[C]//ACL. Morristown, NJ, USA, 2004:178-181
- [9] Longhua Qian, Gougong Zhou, Fang Kong, and Peide Qian. Exploiting constituent dependencies for tree

- kernel-based semantic relation extraction[C]//COLING. Manchester, 2008 :697-704
- [10] Goudong Zhou, Min Zhang, Dong Hong Ji, and QiaoMing ZHU. Tree Kernel-based Relation Extraction with Context-Sensitive Structured Parse Tree Information[C]//EMNLP/CoNLL. Prague, Czech, 2007 :728-736
- [11] Wanxiang Che, Jianmin Jiang, Zhong Su, Yue Pan, and Ting Liu. Improved-Edit-Distance Kernel for Chinese Relation Extraction. IJCNLP. 2005: 132-137
- [12] 庄成龙, 钱龙华, 周国栋. 基于树核函数的实体语义关系抽取方法研究[J]. 中文信息学报. 2009, 23 (1) :3-9.
- [13] Seokhwan Kim and Gary Geunbae Lee. A Graph-based Cross-lingual Projection Approach for Weakly Supervised Relation Extraction[C]//ACL. Jeju, Republic of Korea, 2012:48-53
- [14] David Yarowsky and Grace Ngai. Inducing multilingual POS taggers and NP bracketers via robust projection across aligned corpora[C]//ACL, Pittsburgh, Pennsylvania, 2001:1-8
- [15] David Yarowsky and Grace Ngai, and Richard Wicentowski. Inducing multilingual text analysis tools via robust projection across aligned corpora[C]//ACL, San Diego,2001:1-8
- [16] Paola Merlo, Suzanne Stevenson, Vivian Tsang, and Gianluca Allaria. A multilingual paradigm for automatic verb classification[C]//ACL, Philadelphia, Pennsylvania, 2001:207-214
- [17] Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. Bootstrapping parsers via syntactic projection across parallel texts[J]. Natural Language Engineering, 11 (03) :311-325
- [18] Sebastian Pado and Mirella Lapata. Cross-lingual annotation projection of semantic roles[J]. Journal of Artificial Intelligence Research, 36 (1) :307-340
- [19] Wenjie Li, Peng Zhang, Furu Wei, Yuexian Hou, and Qin Lu. A Novel Feature-based Approach to Chinese Entity Relation Extraction[C]//ACL. Columbus, Ohio, USA. 2008:89-92.
- [20] 车万翔,刘挺,李生. 实体关系自动抽取[J]. 中文信息学报. 2005,19(2):1-6.
- [21] 虞欢欢,钱龙华,周国栋,朱巧明. 基于合一句法和实体语义树的中文语义关系抽取[J]. 中文信息学报. 2010,24(5):17-23.
- [22] 黄瑞红,孙乐,冯元勇,黄云平. 基于核方法的中文实体关系抽取研究[J]. 中文信息学报. 2008,22(5):102-108.
- [23] Min Zhang, Jie Zhang, Jian Su, and Goudong Zhou. A Composite Kernel to Extract Relations between Entities with both Flat and Structured Features[C]//In Proceedings of COLING-ACL. Sydney, Australia. 2006:825-832.
- [24] Guodong Zhou, Longhua Qian, and Jianxi Fan. Tree kernel-based semantic relation extraction with rich syntactic and semantic information[J]. Information Sciences, 2010:1313-1325.