

中文嵌套命名实体关系抽取研究

许浩亮 李雁群 何云琪 钱龙华[†]

苏州大学计算机科学与技术学院, 苏州 215006; [†] 通信作者, E-mail: qianlonghua@suda.edu.cn

摘要 为了解决嵌套命名实体关系抽取研究缺乏相关语料库这一问题, 在现有中文命名实体语料库的基础上, 将人工标注与机器学习相结合来抽取其语义关系。人工标注一个中文嵌套命名实体关系语料库, 然后分别采用支持向量机和卷积神经网络等方法, 进行中文嵌套实体关系抽取实验。实验结果表明, 在人工标注实体的中文嵌套命名实体语料上, 嵌套实体关系抽取的性能非常好, F1 指数达到 95% 以上, 而在自动识别实体上的抽取性能尚不理想。

关键词 嵌套实体关系抽取; 信息抽取; 支持向量机; 卷积神经网络

Research on Chinese Nested Named Entity Relation Extraction

XU Haoliang, LI Yanqun, HE Yunqi, QIAN Longhua[†]

School of Computer Science & Technology, Soochow University, Suzhou 215006;

[†] Corresponding author, E-mail: qianlonghua@suda.edu.cn

Abstract Nested named entities relationship extraction research lacks corresponding benchmark corpora. To solve this problem, manual annotation with machine learning are combined to extract their semantic relationships from an existing Chinese named entity recognition corpus. The authors manually annotate a Chinese nested named entity relation corpus from existing Chinese named entity recognition and conduct experiments with relation extraction between nested named entities via support vector machines (SVM) and convolutional neural network (CNN) models respectively. The experimental results show that the nested entity relation extraction performs excellently on the corpus with manually labeled entities, obtaining an F1 score of over 95%, while it falls short of expectations with automatically recognized entities.

Key words nested entity relation extraction; information extraction; support vector machines; convolutional neural network

信息抽取的目的是从无结构的自由文本中抽取实体及其相互关系, 并转化为结构化表达形式, 从而为知识库的构造提供数据基础^[1-2]。命名实体间语义关系抽取(简称关系抽取)指从文本中提取实体间的语义关系。关系抽取研究对自然语言处理(如问答系统、文档摘要、知识库、本体库等)具有重要的意义。

近十几年来, 得益于基准语料库(如 ACE^[3], SemEval^[4])的出现, 关系抽取的研究取得长足的进步。无论是传统的机器学习方法^[5-6], 还是近年来流行的深度学习方法^[7-10], 关系抽取研究都成为机

器学习算法的试金石。但是, 目前语料库中定义的实体大部分是单一层次的简单实体(如 ACE), 抽取的关系也限于简单实体之间。虽然有一些生物医学领域的嵌套实体语料库^[11], 但是没有定义这些嵌套实体内部的语义关系。另一方面, 由于嵌套实体含有丰富的实体信息及实体间相互关系, 提取这些嵌套实体之间的语义关系有助于丰富知识库的内容。

针对上述情况, 本文在《人民日报》中文实体语料库的基础上, 通过自动生成和手工标注, 构建一个中文嵌套实体语料库, 并进一步标注嵌套实体内部的语义关系, 然后在该语料上分别使用传统的

国家自然科学基金(2017YFB1002101)资助

收稿日期: 2018-04-15; 修回日期: 2018-08-06; 网络出版日期: 2018-08-22

机器学习方法和深度学习方法,进行嵌套实体关系抽取实验。

1 相关工作

目前,关系抽取研究中常用的语料库有 ACE 语料库和 SemEval 2010 语料库。ACE 2005 语料库标注了实体及其相互间语义关系,实体类型分为 7 类(如 PER, ORG, GPE 等),语义关系则有 6 个大类(如 PART-WHOLE, ORG-AFF, GEN-AFF 等),并可细化为 18 个小类。该语料标注的是单一层次的简单实体,语义关系也是发生在简单实体之间。SemEval 2010 标注的都是句子中的名词对,而非实体对之间的语义关系,共有 10 类,其中有 1 个类型不区分关系论元的先后。

目前,没有被广泛认可的中文嵌套命名实体语料库。中文命名实体语料有来源广泛的 MSRA 语料、新闻领域的 1998 年 1 月份《人民日报》语料。由于《人民日报》语料中包含部分嵌套命名实体的标注信息,所以中文嵌套实体识别研究大都基于《人民日报》语料。但是,这些实体标注信息并不完整,存在漏标现象,更没有标注嵌套实体之间的语义关系,如“[中共中央/nt 台湾/ns 工作/vn 办公室/n]nt”转换为嵌套结构“[[[中共中央]nt [台湾]ns 工作办公室]nt”^①,漏标“[中共]nt”这个实体。

传统的关系抽取方法可以分为基于特征向量的方法和基于核函数的方法,前者包括最大熵模型(MaxEnt)^[5]和支持向量机(SVM)^[6],后者有基于依存路径^[12]和基于句法树^[13-15]的两种方法。近年来,神经网络模型在语义关系抽取研究中广泛应用,如卷积神经网络(CNN)^[7-8]模型、双向长短期记忆网络模型(Bi-LSTM)^[9]以及 CNN 和 LSTM 相混合的模型^[10],这些模型能较好地捕获关系实例表达的语义信息。

2 中文嵌套实体关系语料库构建

2.1 中文嵌套实体和实体关系定义

根据命名实体中是否包含其他实体,可以将命名实体分为简单命名实体和嵌套命名实体。简单实体指内部不包含其他实体的实体,其实体层次是单一的,如“[北京]ns”表示一个地名实体。嵌套实体指实体内部嵌套一个或多个命名实体,这种嵌套可以是多层次的,嵌套实体主要存在于地名和机构名

实体中。嵌套在里面的实体称为内部实体,最外层的实体称为外部实体,如外部实体“[[[中共]nt [北京]ns 市委]nt 宣传部]nt”包含“[中共]nt”、“[北京]ns”和“[中共北京市委]nt”等 3 个内部实体。

嵌套实体关系指实体内部嵌套实体之间的语义关系,如嵌套命名实体“[[[中共]nt [北京]ns 市委]nt 宣传部]nt”中,就包含着多个语义关系,即“[中共北京市委宣传部]nt”隶属于“[中共北京市委]nt”,而后者又隶属于“[中共]nt”,且位于“[北京]ns”。从这个例子可以看出,结构复杂的嵌套实体中蕴含丰富的语义关系。

2.2 语料库构建

《人民日报》语料含有部分嵌套实体的标注信息,为了减少标注工作量,本文将其作为构建嵌套命名实体关系语料库的基础,并采用半自动的方式进行标注,具体过程包含两个步骤。

1) 嵌套命名实体的标注。首先自动提取语料中已经标注的部分嵌套实体信息,然后人工标注漏掉的嵌套实体。为了减少重复工作,标注对象是实体而不是一个实体的多个引用。如嵌套实体“[中共/j 北京/ns 市委/n 宣传部/n]nt”经提取后变成两层嵌套实体“[中共 [北京]ns 市委宣传部]nt”,但是这其中还漏掉部分实体,因此还需要人工调整为完整的嵌套实体“[[[中共]nt [北京]ns 市委]nt 宣传部]nt”。

2) 嵌套实体关系的标注。在上述《人民日报》嵌套命名实体语料库的基础上,人工标注嵌套命名实体之间的语义关系。语义关系类型参考 ACE RDC 2005 中文语料的关系类型体系,只不过关系类型数量要少得多。例如嵌套实体“[[[中共]nt [北京]ns 市委]nt 宣传部]nt”中,“中共”和“中共北京市委”存在语义关系“Part-Whole.Subsidiary”,即属于部分整体关系(“Part-Whole”)中的隶属子关系“Subsidiary”。

2.3 语料库一致性检验

为了衡量语料库标注的一致性,我们安排两名志愿者同时进行标注。标注分两个阶段进行,第一阶段,两名志愿者经过初步培训后对语料库进行标注;然后比较他们之间的差异,重新调整标注要求,再进行第二阶段的标注调整,调整结束后计算最终一致性。采用常规的 P , R 和 $F1$ 指数来评估语料库标注的一致性,其中, P 为准备率, R 为召回率, $F1$ 为

① 嵌套实体的类型标注采用《人民日报》语料的格式,即 nr 表示人名, ns 表示地名, nt 表示组织名。

两者的调和平均。

对于嵌套实体标注而言, 第一阶段标注的一致性结果 P , R 和 $F1$ 分别为 92.5%, 93.4% 和 93.0%, 两名志愿者的差异主要体现在对地名的嵌套结构理解不一致; 第二阶段标注的一致性结果 P , R 和 $F1$ 分别为 99.2%, 99.3% 和 99.3%。对于嵌套实体关系标注而言, 第二阶段标注的 P , R 和 $F1$ 值分别为 97.7%, 97.3% 和 97.5%。

2.4 语料库统计

2.4.1 嵌套命名实体统计

表 1 和 2 分别列出标注后的《人民日报》语料中外部实体和内部实体的统计情况, 其中外部实体可进一步分为无嵌套结构和有嵌套结构。内部实体约占所有实体的 14%。

从表 1 和 2 可以看出: 1) 无嵌套结构的外部实体中, 大部分是地名(约 47%)和人名(约 43%), 少量的组织名(约 10%), 如“[中国]ns”、“[邓小平]nr”和“[联合国]nt”; 2) 有嵌套结构的外部实体中, 绝大部

表 1 《人民日报》语料外部实体统计

Table 1 Nested entity statistics of the People's Daily corpus

类型	无嵌套		有嵌套	
	数量	占比/%	数量	占比/%
人名	19981	42.93	0	0
地名	21856	46.96	712	10.41
组织名	4705	10.11	6129	89.59
全部	46542	100.00	6841	100.00

表 2 《人民日报》语料内部嵌套实体统计

Table 2 Internally nested entity statistics of the People's Daily corpus

外部实体类型	人名	地名	组织名	小计
人名	0	0	0	0
地名	22	820	18	859
组织名	87	5605	1994	7685
全部	109	6425	2012	8546

表 3 《人民日报》语料嵌套实体关系统计

Table 3 Nested entity relationships statistics of the People's Daily corpus

类别	数量	占比/%	对应 ACE 2005 类型	含义
Sub	3611	34.09	Part-Whole.Subsidiary	部分与整体的附属关系
Loc	3355	31.67	Gen-Aff.Org-Location	通用附属的组织位置关系
Geo	1060	10.01	Part-Whole.Geo	部分与整体的地理关系
Fou	48	0.45	Org-Aff.Founder	组织附属的创始人关系
None	2519	23.78	—	—
全部	10593	100.00	—	—

分是组织名(约 90%), 少量的地名(约 10%), 如“[[上海市]ns 红十字会]nt”的外部实体“[上海市红十字会]nt”; 3) 内部实体中, 大部分是地名(约 75%), 小部分是组织名(约 24%), 还有极少数人名(约 1%), 如“[[上海市]ns 红十字会]nt”中的“[上海市]ns”。

2.4.2 嵌套实体关系统计

统计《人民日报》语料上嵌套实体关系类型的数量分布情况, 结果见表 3。与 ACE 2005 语料不同, 嵌套实体关系类型只有 4 类。从表 3 可以看出: 1) 嵌套实体中已含丰富的语义关系, 具有实体关系的正例超过 3/4(76%); 2) 嵌套实体关系主要类型为 Sub 和 Loc, 各约占 45% 和 42%, 只有少部分的 Geo(约 13%)和极少部分的 Fou(约 1%)。

3 嵌套实体关系抽取方法

本文采用两种方法来抽取嵌套实体关系: 基于 SVM 的传统机器学习方法以及基于卷积神经网络的深度学习方法。

3.1 基于 SVM 的嵌套实体关系抽取

本文使用如下 4 组特征: 1) 实体字符串, 即两个实体本身的字符串; 2) 实体类型, 即两个实体的命名实体类型; 3) 实体所在层数及其他组合特征, 即实体所在嵌套层数以及两个实体之间的位置关系; 4) 实体的后缀, 即实体结尾的字或词。每组中各个特征的具体说明如表 4 所示。

3.2 基于 CNN 的嵌套实体关系抽取

3.2.1 语料预处理

首先在外实体中插入“[”和“]”来表示内部实体的边界及其嵌套层次关系, 如将“上海医科大学附属中山医院”处理成“[[[上海]医科大学]附属[中山]医院]”, 然后对整个外部实体以“[”和“]”作为分隔符分割成单词序列。例如, 上述实体经分割后, 变成“[₀ [₁ [₂ 上海₃]₄ 医科大学₅]₆ 附属₇ [₈ 中山₉]₁₀ 医院₁₁]₁₂”, 其中单词的下标表示它的序号。

表 4 特征选择
Table 4 Feature selection

特征	特征说明	
实体字符串	eName1	实体 1
	eName2	实体 2
实体类型	(eType1, eType2)	实体 1、实体 2 的命名实体类型的二元组合
	dLayer	实体 1 和实体 2 不在同一层, 所在层数之差
实体所在层数及其他组合特征	Adjacent	实体 1 与实体 2 在同一层, 是否为相邻的实体
	HavingOneEn	实体 1 与实体 2 在同一层, 且实体不相邻, 实体之间是否存在另一个独立的实体
	LastChar1	实体 1 的最后一个字
实体后缀	LastChar2	实体 2 的最后一个字
	LastTwoChars1	实体 1 的最后两个字
	LastTwoChars2	实体 2 的最后两个字

3.2.2 CNN 模型

本文采用经典的卷积神经网络^[16]进行嵌套命名实体关系抽取实验。该网络主要包括输入层、向量表示层、卷积层、池化层和输出层等。输入层包含预处理后的单词序列和两个实体的类型; 向量表示层将单词、单词与两个实体间的相对位置和两个实体的类型分别转换成向量表示, 然后将它们串接起来, 构成实例向量; 在实例向量中, 应用 4 个宽窗口卷积操作得到特征图; 池化层对特征图进行最大池化得到池化向量; 最后, 池化向量经随机丢弃后输入到输出层中的全连接层和 Softmax 层, 从而获得关系类别标签。

3.2.3 模型参数设置

根据以往卷积神经网络上的关系抽取研究^[17-18]以及参数调整过程, 本文实验中 CNN 模型的参数设置如表 5 所示。

4 实验结果与分析

本文采用常规的 P , R 和 $F1$ 指数来评估嵌套实

体关系抽取的性能。采用十折交叉验证方法, 即将《人民日报》嵌套实体关系语料划分为 10 份, 其中 1 份作为测试集, 另外 9 份作为训练集, 总体性能取 10 次结果的平均值。

4.1 人工标注实体下 SVM 模型各种特征对性能的贡献

将 SVM 模型的特征分为 6 组, 基本特征 eName1, eName2, eType1, eType2 构成基准系统 (baseline); 依次加入特征 dLayer, Adjacent, HavingOneEn, LastChar1 以及 LastChar2, LastTwoChar1 和 LastTwoChar2 等, 构成另外 5 组特征组合。

表 6 列出人工标注实体下《人民日报》嵌套语料上各个特征对关系抽取性能的贡献, 其中每一列的最高性能用粗体表示。从表 6 可以看出以下三点。

1) 命名实体的最后单字特征 (LastChar1 和 LastChar2) 对性能的贡献最大, $F1$ 值提高约 7%, 最后双字特征 (LastTwoChar1 和 LastTwoChar2) 对 $F1$ 值提高约 2%, 说明实体的后缀能很好地表示实体的性质及其相互关系。

2) 两实体层差特征 (dLayer) 次之, $F1$ 值提高约

表 5 CNN 模型参数设置
Table 5 Parameter Settings of CNN Model

参数	取值
窗口	2, 3, 4, 5
卷积核	300
丢弃率	0.5
批量大小	32
迭代次数	30
词向量维度	300
位置向量维度	32
实体类型向量维度	16

表 6 SVM 模型上多种特征累加的性能
Table 6 Performance of multiple cumulative features on SVM model

实验分组	$P/\%$	$R/\%$	$F1/\%$
Baseline	76.29	85.72	80.71
+dLayer	80.82	86.14	83.39
+Adjacent	81.76	86.07	83.85
+HavingOneEn	85.66	87.08	86.36
+LastChar1, LastChar2	93.93	93.52	93.72
+LastTwoChar1, LastTwoChar2	95.69	95.31	95.50

3%, 通常直接嵌套(即层差为 1)的两个实体之间存在关系的可能性较大, 因而该特征主要提高准确率。另外, 实体邻近特征(Adjacent)与 dLayer 特征存在冗余, 因此 F1 值略微提高(约 0.5%)。

3) 实体间隔特征(HavingOneEn)也使 F1 值提高约 3%, 即当两个实体之间存在另一个实体时, 这两个实体之间通常不存在关系, 故准确率明显提高。

综上所述, 所有的特征累加起来的实体关系抽取性能最好, F1 值超过 95%, 因此后续 SVM 模型的实验都采用所有特征的累加。

4.2 人工标注实体下 SVM 模型各种关系类型的性能比较

表 7 列出人工标注实体下 SVM 模型实体关系抽取的各个关系类别的性能, 可以看出: 1) 实体关系数量越大, 抽取的性能越高, Sub 类型性能最高(96.53%), Fou 类型性能最低(39.60%); 2) 虽然 Geo 类型的数量是 Loc 类型的 1/3, 但是 F1 降低不到 2%, 说明 Geo 类型容易识别, 因为它表示的是两个地名之间的部分整体关系, 表达模式较单一, 而 Geo 类型表达组织名与地名的位置关系, 略微复杂。

4.3 自动识别实体下 SVM 模型各个关系类型的性能比较

表 8 比较采用嵌套实体自动识别下的关系抽取中各个关系类型的性能。本文采用由内而外的层次模型方法^[19]来识别嵌套命名实体, 即用多个 CRF

模型, 由内到外地识别不同嵌套层次的实体。实验表明, 该方法的嵌套命名实体识别总体性能 P , R 和 $F1$ 为 94.2%, 84.6%和 89.1%, 但内部实体的识别性能 P , R 和 $F1$ 却只有 76.4%, 84.6%和 80.3%。

从表 8 可以看出: 1) 与表 7 中人工标注实体上的关系抽取性能相比, 自动识别实体下关系抽取 F1 降低幅度非常大(约 22%), 说明嵌套实体识别性能是制约嵌套关系抽取性能的主要因素; 2) 与表 7 中人工标注实体上的关系抽取性能类似, 关系类别数量越多, 对应关系抽取的性能就越高, 但由于实体自动识别错误的拖累, Geo 类型的 F1 性能下降约 37%。

对其中一个测试集的关系抽取错误样例进行分析, 发现嵌套实体识别的结果严重地影响后续实体关系的抽取。原因有两方面: 1) 嵌套实体识别产生的假负例导致关系抽取的假负例较多(占有所有正例的 62%), 如在实体关系“(中纪委, 中纪委监察部, Sub)”中, “中纪委”和“中纪委监察部”这两个实体都没识别出来; 2) 在识别出的嵌套实体关系中, 33%是假正例, 如实体关系“(湖北省, 科委湖, Geo)”中的“科委湖”是假实体, 因此该实体关系也是假正例。

4.4 人工标注实体下 CNN 模型嵌套关系抽取性能

表 9 列出 CNN 模型在人工标注实体下各个关系类别的性能, 可以看出以下两点。

表 7 人工标注实体下 SVM 模型嵌套关系抽取性能

Table 7 Performance of nested relation extraction of SVM model under manually annotated entities

关系类型	数量	占比%	$P\%$	$R\%$	$F1\%$
Sub	3611	44.71	96.81	96.24	96.53
Loc	3355	41.57	95.08	94.91	94.98
Geo	1060	13.13	93.40	92.94	93.06
Fou	48	0.59	39.23	40.00	39.60
全部	8074	100.00	95.69	95.31	95.50

表 8 自动识别实体下 SVM 模型嵌套关系抽取性能

Table 8 Performance of nested relation extraction on SVM model under automatic recognition entities

关系类型	数量	百分比	$P\%$	$R\%$	$F1\%$
Sub	3611	44.71	84.70	69.87	76.51
Loc	3355	41.57	79.49	69.55	74.14
Geo	1060	13.13	66.87	51.03	57.77
Fou	48	0.59	40.00	33.06	35.71
全部	8074	100.00	80.49	67.38	73.33

表 9 人工标注实体下 CNN 模型嵌套关系抽取性能
Table 9 Performance of nested relation extraction of CNN model under manually annotated entities

关系类型	数量	占比/%	P/%	R/%	F1/%
Sub	3611	44.71	95.25	91.68	93.41
Loc	3355	41.57	90.26	94.90	92.50
Geo	1060	13.13	91.60	94.96	93.14
Fou	48	0.59	40.00	35.28	37.38
全部	8074	100.00	92.61	93.47	93.04

1) 与表 7 相比, CNN 模型的嵌套实体关系抽取 F1 值比 SVM 模型降低约 2.5%。这主要是由于嵌套实体关系抽取中结构化特征起重要作用, 而 CNN 中没有 SVM 中的显式结构特征。例如, 在嵌套实体“[[[上海]医科大学]附属[中山]医院]”中, CNN 模型能够学到局部特征, 但不能获取全局结构特征。

2) 与表 7 不同, 除数量特别少的 Fou 类别外, 另外 3 个类别虽然数量相差较大, 但是性能相差不大(1%左右)。这说明 CNN 只需要较少的训练数据, 就可以挖掘出隐含的语义特征, 从而生成鲁棒性较好的模型。

虽然 CNN 模型的嵌套实体关系抽取性能比 SVM 模型 F1 降低约 2.5%, 但是, SVM 模型需要通过大量人工提取的特征来提高性能, 而 CNN 模型的特征不需要人工提取特征, 能够更好地刻画数据的内在信息。因此, 对于嵌套命名实体关系抽取, CNN 模型具有一定的潜力。

5 结语

中文嵌套实体关系语料库的构建为嵌套实体识别和关系抽取提供了一个基准平台, 有助于信息抽取领域进一步的研究。本文在已有的中文命名实体语料的基础上, 人工标注中文嵌套实体关系语料, 并利用 SVM 模型和深度学习模型, 抽取中文嵌套实体之间的语义关系。实验表明, 中文嵌套命名实体关系抽取在人工标注实体上取得很好的性能, 但是在自动识别实体上的性能却不尽如人意, 这主要是由中文嵌套命名实体识别性能不高导致。另外, CNN 模型的性能虽然比 SVM 模型略低, 但是对于嵌套命名实体关系抽取也很有效。

今后的工作包括两个方面: 1) 考虑使用深度学习模型来提高嵌套命名实体识别的性能; 2) 考虑嵌套命名实体识别及其关系抽取的联合模型, 从而同时提高两者的性能。

参考文献

- [1] Zheng Suncong, Wang Feng, Bao Hongyun, et al. Joint extraction of entities and relations based on a novel tagging scheme [EB/OL]. (2017-06-07)[2017-06-29]. <https://arxiv.org/pdf/1706.05075.pdf>
- [2] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition [EB/OL]. (2016-04-07)[2017-06-29]. <https://arxiv.org/pdf/1603.01360.pdf>
- [3] Qian Longhua, Zhou Guodong, Kong Fang, et al. Semi-supervised learning for semantic relation classification using stratified sampling strategy // Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, 2009: 1437-1445
- [4] Hendrickx I, Su N K, Kozareva Z, et al. SemEval-2010 task 8: multi-way classification of semantic relations between pairs of nominals // The Workshop on Semantic Evaluations: Recent Achievements and Future Directions. Uppsala: Association for Computational Linguistics, 2009: 94-99
- [5] Kambhatla, Nanda. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations // ACL 2004 on Interactive Poster and Demonstration Sessions. Barcelona: Association for Computational Linguistics, 2004: 22
- [6] Zhou Guodong, Su Jian, Zhang Jie, et al. Exploring various knowledge in relation extraction // Proceedings of the 43rd Annual Meeting of the ACL. Ann Arbor, 2005: 427-434
- [7] Nguyen T H, Grishman R. Relation extraction: perspective from convolutional neural networks // The Workshop on Vector Space Modeling for Natural Language Processing. Denver, 2015: 39-48
- [8] Wang Linlin, Cao Zhu, Melo G D, et al. Relation classification via multi-level attention cnns // Meeting of the Association for Computational Linguistics.

- Berlin, 2016: 1298–1307
- [9] Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures // Meeting of the Association for Computational Linguistics. Berlin, 2016: 1105–1116
- [10] Zheng Suncong, Xu Jiaming, Zhou Peng, et al. A neural network frame work for relation extraction: learning entity semantic and relation pattern. Knowledge-Based Systems, 2016, 114(1): 12–23
- [11] Ohta T, Tateisi Y, Kim J D. The GENIA corpus: an annotated research abstract corpus in molecular biology domain // International Conference on Human Language Technology Research. San Diego, 2002: 82–86
- [12] Culotta A, Sorensen J. Dependency tree kernels for relation extraction // Meeting of the Association for Computational Linguistics. Barcelona: DBLP, 2004: 423–429
- [13] Qian Longhua, Zhou Guodong, Kong Fang, et al. Exploiting constituent dependencies for tree kernel-based semantic relation extraction // COLING 2008, International Conference on Computational Linguistics Manchester: DBLP, 2008: 697–704
- [14] Zhang Min, Zhang Jie, Su Jian, et al. A Composite kernel to extract relations between entities with both flat and structured features // International Conference on Computational Linguistics and Meeting of the Association for Computational Linguistics. Sydney: DBLP, 2006
- [15] Zhou Guodong, Zhang Min, Ji D H, et al. Tree kernel-based relation extraction with context-sensitive structured parse tree information // Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Prague: DBLP, 2007: 728–736
- [16] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network // The 25th International Conference on Computational Linguistics. Dublin, 2014: 2335–2344
- [17] Qin Pengda, Xu Weiran, Guo Jun. An empirical convolutional neural network approach for semantic relation classification. Neurocomputing, 2016, 190: 1–9
- [18] Santos C N D, Xiang Bing, Zhou Bowen. Classifying relations by ranking with convolutional neural networks. Computer Science, 2015, 86(86): 132–137
- [19] Alex B, Haddow B, Grover C. Recognising nested named entities in biomedical text // The Workshop on Bionlp 2007: Biological, Translational, and Clinical Language Processing. Prague: Association for Computational Linguistics, 2007: 65–72