

文章编号: 1003-0077(2009)01-0003-06

基于树核函数的实体语义关系抽取方法研究

庄成龙, 钱龙华, 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)

摘 要: 该文描述了一种改进的基于树核函数的实体语义关系抽取方法, 通过在原有关系实例的结构化信息中加入实体语义信息和去除冗余信息的方法来提高关系抽取的性能。该方法在最短路径包含树的基础上, 首先加入实体类型、引用类型等与实体相关的语义信息, 然后对树进行裁剪, 去掉修饰语冗余和并列冗余信息, 并扩充所有格结构, 最后生成实体语义关系实例。在 ACE RDC 2004 基准语料上进行的检测实验和 7 个关系大类抽取的实验表明, 该方法在较大程度上提高了实体语义关系识别和分类的效果, F 值分别达到了 79.1% 和 71.9%。

关键词: 计算机应用; 中文信息处理; 实体关系抽取; 树核函数; 语义信息

中图分类号: TP391

文献标识码: A

Research on Tree Kernel-Based Entity Semantic Relation Extraction

ZHUANG Cheng-long, QIAN Long-hua, ZHOU Guo-dong

(School of Computer Science & Technology, Soochow University, Suzhou, Jiangsu 215006, China;
Jiangsu Provincial Key Lab. of Computer Information Processing Technology, Suzhou, Jiangsu 215006, China)

Abstract: This paper describes an improved tree kernel-based approach to entity semantic relation extraction, where the performance is improved by incorporation of entity-related semantic information into, the structured representation of relation instances and the pruning of redundant information. Starting from the Shortest Path-enclosed Tree for a relation instance, entity-relation semantic information, such as entity types, subtypes, and mention types etc., are first uniformly appended. Then modifications to noun phrases and redundant information in conjunction coordination structures are removed away, but the possessive structure is further included. With such generated appropriate representation of the relation instance, experiments on the ACE RDC 2004 benchmark corpus shows that our method significantly improves the performance, achieving the F-measure of 79.1% and 71.9% on the task of relation detection and top-level relation extraction respectively.

Key words: computer application; Chinese information processing; entity relation extraction; tree kernel function; semantic information;

1 引言

进入 21 世纪以来, 随着科技的不断进步, 尤其是互联网技术的快速发展, 现实世界中的信息量迅猛增加, 远远超出了人类阅读的能力。如何过滤无用信息并从中抽取人们所需要的特定信

息成为一个迫切需要解决的难点。信息抽取的主要目的就是从小结构的自然语言文本中抽取特定的事件、事实等信息, 再转化为结构化或半结构化的信息, 然后储存在数据库中, 供查询以及进一步分析利用。信息抽取最早是在美国国防高级研究计划局 (DARPA) 资助的信息理解会议 (MUC, 1987-1998)^[1] 上提出, 并逐渐发展成为自然语言

收稿日期: 2008-07-07 定稿日期: 2008-09-01

基金项目: 国家 863 高技术研究发展计划资助项目 (2006AA01Z147); 国家自然科学基金资助项目 (60673041)

作者简介: 庄成龙 (1985—), 男, 硕士生, 主要研究方向为信息抽取; 钱龙华 (1966—), 男, 副教授, 硕导, 在职博士生, 主要研究方向为自然语言处理; 周国栋 (1967—), 男, 教授, 博导, 主要研究方向为自然语言处理。

处理(NLP)领域的一个重要分支。MUC会议停止后,由美国国家标准技术局(NIST)资助的“自动内容抽取”(ACE)评测会议^[2]进一步推动着信息抽取研究的发展。

ACE中信息抽取的任务主要包括实体识别和跟踪(EDT, Entity Detection and Tracking)、关系识别和描述(RDC, Relation Detection and Characterization)以及事件识别和描述(EDC, Event Detection and Characterization)等。本文的研究重点是实体关系的识别与描述,即实体语义关系的抽取,其目的是从文本中找出实体对之间的语义关系并对它进行分类。例如“比尔·盖茨是微软公司的总裁”中包含了一种“雇佣”(EMP-ORG)关系,表示实体“比尔·盖茨”(PER)受雇于实体“微软公司”(ORG)。实体关系的抽取在实际应用中的范围很广,对于信息抽取、问答系统、机器翻译等领域的发展都有着重要作用。

本文使用卷积树核方法^[3]来进行关系抽取,它通过直接计算两个实体关系对象(即句法树)的相同子树的个数来比较相似度。由于核方法可以充分利用特征方法无法表示的结构化信息,因此近年来越来越多的研究人员开始研究和使用该方法,例如:Zelenko et al.^[4], Culotta^[5], Bunescu^[6]和Zhang^[7]。但是以往的研究工作所采用的关系实例表达方式存在着很多冗余信息,从而影响了性能的提高。为了改进关系实例表达的方式,进一步研究结构化信息在实体关系抽取中的作用,我们在文献^[7]的基础上提出了一种优化的树核方法,应用树的修剪策略,在减少冗余信息的同时扩充了原有的树结构,使之包含更丰富的实体语义信息。在ACE 2004基准语料库上的测试表明,该方法能显著提高关系抽取系统的性能。

本文的后续组织结构如下:第二部分回顾了实体关系抽取领域的相关工作;第三部分论述了本文所使用的方法以及树裁剪的策略;第四部分为实验结果和性能分析;最后是全文总结和将来工作的方向。

2 相关工作

现阶段的实体语义关系抽取主要有两种方法:基于特征向量的机器学习方法和基于核函数的机器学习方法。

基于特征向量的学习方法首先需要构造符合特征向量形式的训练数据,然后使用各种机器学习算

法,如支持向量机(SVM)、Winnow等作为学习器构造分类器进行训练和测试。在关系抽取中,典型的基于特征向量的方法包括最大熵模型(Max-Ent)^[8]和支持向量机(SVM)^[9-10]。基于特征向量的关系抽取其研究重点在于如何获取各种有效的词汇、语法和语义等特征,并把它们有效地集成起来,产生描述对象的各种局部特征和简单的全局特征。例如,文献^[10]集成了各种词汇、语法解析树和依存树等特征,同时加入了WordNet和Name List等语义信息,在ACE 2004的语料库上的7个大类的关系抽取中,F指数达到了70.1。但是由于实体间语义关系表达的复杂性和可变性,要进一步找出有效的特征从而提高抽取性能则变成非常困难的问题。

与基于特征向量的方法不同,基于核函数的方法不需要构造高维特征向量空间。核函数方法以结构树为处理对象,通过直接计算两个离散对象(如语法结构树)之间的相似度来进行分类,这使得基于核函数的方法理论上可探索隐含的高维特征空间,从而可以有效地利用句法树中的结构化信息。Zelenko^[4]最早把核函数的方法引入了关系抽取领域,首先在文本的浅层解析树的基础上定义了核函数,并设计了一个用于计算核函数的动态规划算法,然后通过支持向量机(SVM)和表决感知器(Voted Perceptron)等分类算法来抽取实体语义关系,在200篇来自新闻机构(如美联社、《华尔街日报》等)的新闻文章中进行测试,取得了较好的效果。Culotta^[5]通过一些转换规则(如主语依存于谓语、形容词依存于它们所修饰的名词等)将包含两个实体的解析树转换成依存树,并在树节点上增加词性、实体类型、词组块、WordNet上位词等特征,然后定义了基于依存树的核函数并使用SVM分类器进行关系抽取,在ACE RDC 2003基准数据上的5个关系大类的抽取中F指数取得了45.8。Bunescu等^[6]进一步提出了基于最短路径依存树的核函数,通过计算机在依存中两个实体之间的最短路径上的相同节点的数目比较实体关系相似度,虽然在ACE RDC2003基准数据中F指数取得了52.5,但是其召回率较低。Zhang等^[7]设计了一种复合卷积树核函数来进行关系抽取,该方法将卷积树核函数和线性核函数(与实体属性相关,如实体类型、引用类型等)结合起来,充分考虑了影响语义关系的平面特征和结构特征,在ACE 2003和ACE 2004基准数据上的大类关系抽取中F指数分别达到了70.9%和72.1%。

以上研究工作在使用核函数进行关系抽取方面进行了有益的尝试,并取得了较好的性能,但是他们所采用的句法树内部仍然存在着一定的冗余结构信息,同时也没有考虑与实体相关的语义信息,因而性能的提高也遇到了瓶颈。本文采用 Zhang^[7]提出的方法,并在此基础上利用裁剪策略对生成树重新进行改进,同时加入了一些实体语义信息,有效地丰富了关系实例的结构化信息,使得关系抽取的性能得到明显提高。

3 基于树核函数的关系抽取

这部分介绍了本文所使用的卷积树核函数、实验所使用的语料,然后描述了一种新的句法分析树裁剪策略,即如何去除冗余结构以及如何加入实体语义信息的方法等。

3.1 卷积树核函数

卷积树核函数最初由 Collins^[3]引入到自然语言处理领域,该方法是通过计算两棵解析树之间的相同子树的数量来比较解析树之间的相似度。例如有两棵解析树 T1 和 T2,它们之间的相似度可由下列公式来计算:

$$K_c(T_1, T_2) = \sum_{n_1 \in N_1, n_2 \in N_2} \Delta(n_1, n_2)$$

其中 N_j 是 T_j 的节点集合, $\Delta(n_1, n_2)$ 计算以 n_1 和 n_2 为根的共同子树个数,它可以按照下面递归的方法进行计算:

- (1) 如果 n_1 和 n_2 节点处的产生式不同,则 $\Delta(n_1, n_2) = 0$, 否则转向(2);
- (2) 如果 n_1 和 n_2 都是叶子前的一个节点,则 $\Delta(n_1, n_2) = 1 \times \lambda$, 否则转向(3);
- (3) 递归地计算 $\Delta(n_1, n_2)$:

$$\Delta(n_1, n_2) = \lambda \prod_{k=1}^{\#ch(n_1)} (1 + \Delta(ch(n_1, k), ch(n_2, k)))$$

其中 $\#ch(n_1)$ 是节点 n 的孩子节点数目, $ch(n, k)$ 是节点 n 的第 k 个孩子节点, $\lambda (0 < \lambda < 1)$ 是衰退因子。卷积树核函数计算的时间复杂度为 $O(|N_1| \cdot |N_2|)$ 。

3.2 实验语料库

本文实验数据采用的是 ACE RDC 2004 基准语料库。ACE 2004 标注语料库包含了从广播、新闻、报纸等收集的各种新闻报道。ACE RDC 任务中预定义了 7 个关系大类和 23 个关系子类。为了便于比较,我们将关系抽取的任务限定在 ACE

2004 所定义的 7 个大类,详见表 1。

表 1 ACE RDC 2004 所定义的 7 个大类的语义关系

关系类别	描述内容
PHYS	物理关系,描述了实体之间物理上的临近关系
PER-SOC	人物社会关系,描述了人们个体之间的社会关系,拥有这类关系的两个实体必须是人的实体
EMP-ORG	雇员组织机构关系,描述了雇员与组织机构的雇佣关系;公司成员与公司或者子公司与总公司的关系
ART	描述了施事者对于人造物品的拥有或者发明制造等关系
OTHER-AFF	描述了未被 PER-SOC 关系涵盖的人们个体之间或者个体与团体之间的其他的关系
GPE-AFF	描述了个人和组织以及行政区域之间的关系
DISC	两个实体之间存在的整体和部分或者成员的关系

在处理基准语料库时,首先从标注文件中提取出所需的实体信息,包括实体的类型、引用类型、实体中心词等,然后从文本文件中提取出文本信息进行分句和句法分析,本文采用 Charniak 句法分析器^[11]生成句法分析树。

3.3 关系实例的生成

在基于树核函数的关系抽取中,关键问题是如何表示实体关系的结构化信息,即抽取句法树中的哪些部分作为关系实例的表达方式。Zhang^[7]提出了五种句法树的抽取方法,其中最短路径包含树(SPT, Shortest Path-enclosed Tree)取得的效果最好,它的生成方法是以句法树中连接两个实体的最短路径为边界,保留路径及其下面所包含的所有信息,同时去除路径外的所有信息。

在实际情况中,即使按照路径包含树来裁剪句法树,留下的 SPT 还是包含了许多冗余信息,并且在树的裁剪过程中也把一些有用的上下文信息删除了,这些都在一定程度上降低了关系抽取的效果。本文通过加入实体语义信息和进一步消除冗余信息等途径来进一步提高关系抽取的性能:

第一,增加与实体相关的语义特征。通常,实体关系与实体的语义属性密切相关,例如 PER-SOC 关系描述了个体之间的社会关系,因而它的两个实体必定是 Person 类型。在基于特征的方法中,实体本身的属性或者属性组合是构造向量的一个重要特

征。在实验中我们加入了与实体相关的语义信息,如实体大类 TP、子类 ST、引用类型 MT 等,从而生成一棵语义信息扩展树 SEPT (Semantic Extended PT),如图 1 所示。根据卷积核函数的原理,计算树的相似度时,衰退因子的作用会使得层次越深的节点对整体相似度的贡献越小,因此我们把这些语义信息添加在关系实例树的根节点上。

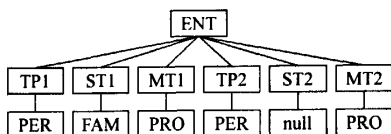


图 1 实体语义信息扩展树 SEPT

第二,消除结构冗余信息。通过对分错的关系

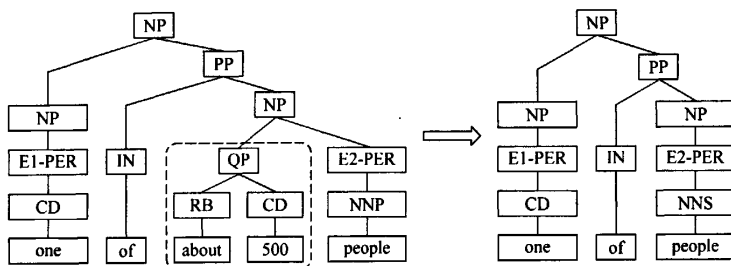


图 2 去除修饰语冗余树 MRPT

并列冗余是指语句中的并列结构冗余。例如在短语“Presidents(E1) of China, Russia(E2) and America……”中,可以清晰地判别出实体 E1 和实体 China、Russian 和 American 之间分别存在着雇佣关系,可事实上分类器很难识别出 E1 和 E2 之间的雇佣关系,究其原因实体 E2 前面的并列部分作为噪音干扰了分类器的正常识别。为了解决这个问

实例的仔细观察,我们发现很多关系无法正确识别的原因是实体关系中存在冗余结构,如修饰语结构、并列结构等等。

修饰语冗余是指实体之前存在的修饰语如冠词、形容词、同位语结构等,这些修饰词对确定实体之间的语义关系几乎没有任何作用,因而在 SPT 树结构中作为噪声影响了分类器的性能。例如:实体“one of about 500 people”和“one of people”生成的树结构在比较相似度的时候就可能会被误认为不相似,因此在生成这类关系实例时通过编写规则,去掉实体对间的多余修饰语,生成去除修饰语冗余树 (MRPT, Modification Removed PT),提高两棵句法树的相似度,进而提高关系识别的概率,如图 2 所示。

题,在识别 E1 和 E2 之间的语义关系是,可以去除并列结构中与实际无关的并列冗余部分。如图 3 所示,在对 (E1, E2) 进行关系识别的时候,把原来的短语转换成“Presidents(E1) of America(E2)”,这样分类器就可以正确地识别出它们之间的关系。我们把这种裁剪掉并列冗余部分的句法树叫 CRPT (Conjunction Removed PT)。

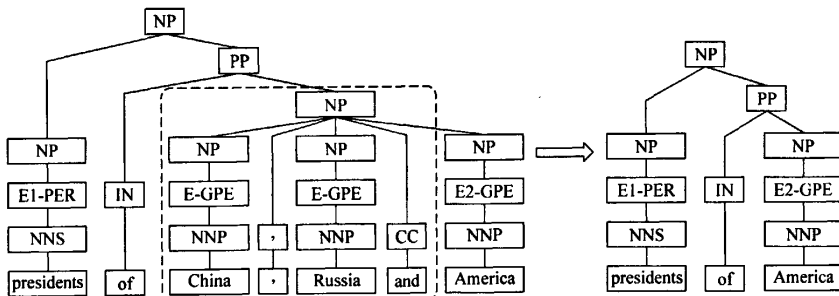


图 3 并列冗余信息的消除 CRPT

最后我们通过实验统计发现,语料库中很多上下文信息对关系识别的效果也有一定的影响,其中比较明显的是短语中的所有格结构。例如在短语

“member of missouri’s supreme”中,实体“member”和“supreme”有着“雇佣组织 (EMP-ORG)”关系,但后者被所有格结构“missouri’s”修饰,因而实

体“member”和实体“missouri”之间不存在任何关系。如果按照路径包含树(PT)的生成方式,抽取出实体“member”和实体“missouri”之间的路径包含树,则生成的实例就是“member of missouri”,从而导致分类器把它们误识别为雇佣关系。出现这种情况的根本原因在于,在生成关系实例时省略了所有格

后面的中心词,因此我们在生成实例时就在第二个实体后面扩充可能存在的所有格结构,一般情况下只要保留所有格的标志词(即’s)即可。这种树称为所有格扩充树(PEPT, Possessive Extended PT),如图 4 所示。

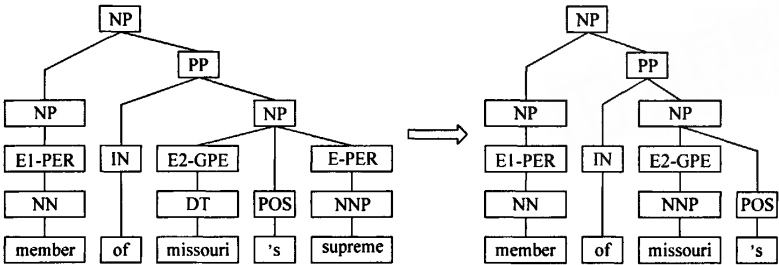


图 4 所有格扩充树 PEPT

通过以上几种方法对关系实例添加语义信息并进行相应裁剪后,最终生成了关系实例树,接下来就是通过实验来验证该方法的有效性。

4 实验结果与分析

本文的实验工具使用基于卷积树核函数^[3]的 SVM 分类器 Tree Toolkits^[12],这是由于卷积树核能够有效地捕获关系实例的结构化信息,并且 SVM 分类器是目前性能最好的分类器之一。实验数据取自 ACE RDC 2004 中的 347 篇(BNEWS/NWIRE)新闻报道,共有 4 307 个关系实例,对 ACE 所定义的 7 个大类进行关系抽取实验。

4.1 实体语义信息对关系抽取的影响

首先在 Zhang^[7]的最短路径树的基础上加入各种实体语义信息,包括实体顺序、大类类型、子类类型和引用类型等。表 2 列出了这些信息对实体关系抽取的影响。实验结果显示:

- 1) 加入实体大类类型信息后 F 值提高了大约 11,表明实体大类类型对关系抽取的贡献非常显著,这说明语义关系受到实体类型的限制,某些关系所属的实体类型是特定的。加入子类型和引用类型后性能也有不同程度的提高,表明子类型和引用类型对关系抽取性能提高也有较大的作用。
- 2) 加入实体类别、GPE 角色、中心词和 LDC 引用类型后性能不但没有提高,反而有所降低。这说明这些属性对关系抽取没有积极作用,可能是由于

有些属性(如实体类别)过于笼统从而不具备区分度,而有些属性(如中心词)又过于分散从而失去泛化能力。

表 2 不同实体语义信息对关系大类抽取性能的影响

	实体属性	P/%	R/%	F/%
1	SPT(baseline)	66.7	50.3	57.4
2	+entity type*	75.7	61.4	67.6
3	+entity subtype*	77.6	62.4	69.1
4	+mention level*	79.1	63.6	70.5
5	+entity class ⁽⁻⁾	78.9	62.7	69.9
6	+GPE role ⁽⁻⁾	78.9	62.4	69.7
7	+head word ⁽⁻⁾	80.8	60.3	69.1
8	+LDC type ⁽⁻⁾	60.4	60.3	68.9

(* 表示该属性起正作用, (-) 起负作用)

4.2 不同生成树对系统性能的影响

表 3 列出了不同的关系实例生成树结构在 ACE RDC 2004 语料库上的关系检测和 7 个大类关系抽取的性能。

通过比较我们发现语义信息扩展树的 F 值性能提高了约 13,效果较明显。这是说明有针对性地加入实体语义信息对关系抽取有较大帮助。在此基础上,加入其他的三种树结构也有不同程度的提高,但是效果相对不是很明显。这是因为一方面名词所有格结构在语料库中的数量相对较少,因而对结果

影响较小。另一方面由于语料库中的句法生成树较复杂,随着生成树的深度不断增加,即使对这些冗余信息进行裁剪,受核函数衰减因子的影响,在比较相似度时作用也削弱了,因而它们对最终效果的影响也变得较小。

表 3 四种不同树在 ACE RDC 2004 数据上测试的性能比较

实例结构	关 系 检 测			7 个大类上的关系抽取		
	准确率(P)/%	召回率(R)/%	F1 值(F1)/%	准确率(P)/%	召回率(R)/%	F1 值(F1)/%
原型系统	84.4	63.6	72.5	66.7	50.3	57.4
+SEPT	87.3	70.3	77.9	78.6	63.4	70.5
+PEPT	87.6	71.2	78.6	78.4	65.5	71.4
+MRPT	87.8	71.2	78.6	78.9	65.5	71.6
+CRPT	86.8	72.6	79.1	78.6	66.9	71.9

表 4 与其他系统性能相比较(ACE RDC 2004)

系 统	关 系 检 测			大类关系抽取		
	P/%	R/%	F1/%	P/%	R/%	F1/%
本系统(基于单树核)	86.8	72.6	79.1	78.6	66.1	71.9
Zhang ^[7] (基于单树核)	—	—	—	74.1	62.4	67.7
Zhao ^[9] (基于特征)	—	—	—	69.2	70.5	70.3
Zhou ^[10] (基于特征)	89.0	66.6	76.2	82.8	62.1	71.0

4.3 总体性能与其他同类系统的比较

在表 4 中把我们的实体关系抽取结果和其他三种关系抽取系统进行了比较,在大类的抽取效果方面比 Zhang^[7]的单树核方法 F 值提高了 4,性能提高较明显,这都得益于我们的改进策略能够更加有效地利用实体语义信息,并且和基于特征的方法比较,比 Zhao^[9]F 值性能提高了 1.6,与基于特征的最好结果 Zhou^[10]相比 F 值也提高了 0.9,这说明核方法与传统的基于特征方法相比,在关系抽取方面具有更好的发展潜力。

5 总结

本文利用卷积树核函数方法进行实体语义关系的抽取,针对目前的方法在关系实例表达方式上存在的问题,采用一种改进的生成关系实例的策略,对原有的实体关系树进行裁剪并且加入实体语义信息,在 ACE RDC 2004 语料库上进行的实验表明,该方法能有效地提高关系抽取的性能。实验中我们发现实体相关语义信息特征对关系抽取的影响最大,加入这些信息后可以在很大程度上改善关系

抽取的性能。同时对关系实例进行除噪处理后,对关系抽取的提高也有一定的效果。复杂结构的关系识别一直是关系抽取中的难题,因此下一步的工作设想是采用一些篇章理论把复杂结构的关系实例简化成可行的形式,提高这些关系实例的抽取性能。

参考文献:

[1] MUC[EB/OL]. http://www.itl.nist.gov/iaui/874.02/related_project/muc/,1987-1998.

[2] ACE. The Automatic Context Extraction Project[EB/OL]. <http://www.ldc.upen.edu/Project/ACE>, 2002-2005.

[3] Collins M,Duffy N. Convolution Kernels for Natural Language[C]//NIPS, 2001.

[4] Zelenko D,Aone C,Richardella A. Kernel Methods for Relation Extraction[J]. Journal of Machine Learning Research,2003,(2): 1083-1106.

[5] Culotta A,Sorensen J. Dependency tree kernels for relation extraction[C]//ACL, 2004: 423-429.

(下转第 34 页)

next 链表中的节点进行了较好的分类,减少了 next 链表的平均长度,且在匹配过程中不再需要比较原 WM 算法中相应 next 链表中的所有节点,能比 WM、OWM 更早地结束循环,所以提高了匹配的效率。

参考文献:

- [1] A. V. Aho and M. J. Corasick. Efficient string matching, an aid to bibliographic search[C]//Communications of the ACM, 1975, 18(6): 333-340
 - [2] B. Commentz_Walter. A string matching algorithm fast on average[R]. Proceedings of the 6th International colloquium on Automata, Languages and Programming, number 71 in Lecture Notes in Computer Science, Springer_Verlag, 1974: 118-132.
 - [3] S. Wu and V. Manber. A fast algorithm for multi_pattern searching[R]. Report TR-94-17, Department of Computer Science. University of Arizona, Tucson, AZ, 1994.
 - [4] Gonzalo Navarro and Mathieu Raffinot. Flexible Pattern Matching in strings. 柔性字符串匹配[M]. 中国科学院计算技术研究所网络信息安全研究组译. 北京: 电子工业出版社 2007. 3 ISBN 978-7-121-03858-7.
 - [5] R. S. Boyer and J. S. Moore. A fast string searching algorithm[C]//Communications of the ACM, 1977, 20(10): 762-772,.
 - [6] 张鑫,等. 一种改进的 Wu-Manber 多关键词算法[J]. 计算机应用, 2003, 23(7): 29-31.
 - [7] 孙晓山,等. 一种改进的 Wu-Manber 多模式匹配算法及应用[J]. 中文信息学报, 2006, 20(2): 47-52.
 - [8] 王素琴, 邹旭楷. 一种优化的并行汉字/字符串匹配算法[J]. 中文信息学报, 1995, 9(1): 49-53.
 - [9] 陈开渠,等. 快速中文字符串模糊匹配算法[J]. 中文信息学报, 2004, 18(2): 58-65.
-
- (上接第 8 页)
- [6] Bunescu R. C. and Mooney R. J. 2005. A Shortest Path Dependency Kernel for Relation Extraction[J]. EMNLP-2005: 724-731.
 - [7] Zhang M, Zhang J, Su J, et al. A Composite Kernel to Extract Relations between Entities with both Flat and Structured Features[C]//ACL, 2006: 825-832.
 - [8] Kambhatla N. Combining lexical, syntactic and semantic features with Maximum Entropy models for extracting relations[C]//ACL(poster), 2004: 178-181.
 - [9] Zhao S B, Grishman R. Extracting relations with integrated information using kernel methods [C]//ACL, 2005: 419-426.
 - [10] Zhou G D, Su J, Zhang J, Zhang M. Exploring various knowledge in relation extraction[C]//ACL, 2005: 427-434.
 - [11] Charniak, Eugene. Intermediate-head Parsing for Language Models[C]//ACL, 2001: 116-123.
 - [12] Moschitti A. A study on Convolution Kernels for Shallow Semantic Parsing[C]//ACL, 2004.