

doi: 10.7690/bgzd.2013.09.017

## 基于机器学习的实体关系抽取方法

刘方驰, 钟志农, 雷霆, 吴烨

(国防科学技术大学电子科学与工程学院, 长沙 410073)

**摘要:** 实体关系抽取是信息抽取的一项重要内容, 总结现有的方法对于该领域的发展具有指导和借鉴意义。结合当前的研究进展, 分析和比较了有监督、无监督和弱监督 3 类关系抽取方法的原理和代表性算法, 总结了各类方法的特性并对关系抽取的发展趋势进行了展望。

**关键词:** 实体关系抽取; 机器学习; 有监督; 无监督; 弱监督

**中图分类号:** TP303 **文献标志码:** A

## Entity Relation Extraction Method Based on Machine Learning

Liu Fangchi, Zhong Zhinong, Lei Lin, Wu Ye

(School of Electronic Science & Engineering, National University of Defense Technology, Changsha 410073, China)

**Abstract:** Relation extraction is an important section of information extraction, summarization of the existing methods is instructional for the developing of this field. Combined with the current research status, firstly, analyzed and compared the principle and representative algorithms of three relation extraction methods, including supervised, unsupervised and semi-supervised based on machine learning, then summarized the characteristic of all the three methods. Finally, put forward outlook for development tendency.

**Key words:** entity relation extraction; machine learning; supervised; unsupervised; semi-supervised

### 0 引言

面对着来自互联网和其他渠道的文本数据的爆炸式增长, 如何从这些非结构化数据中获取所需要的信息成为困扰人们的一个难题, 在这种情况下, 信息抽取技术应运而生。信息抽取 (information extraction) 技术, 是指从一段文本或其他形式的非结构化数据中抽取特定的实体、事件、关系等信息, 形成结构化的数据存储成关系数据库、XML 数据形式, 供用户查询和使用的过程。信息抽取的任务包括命名实体识别、句法分析、关系抽取、篇章分析与推理、知识获取等, 其中命名实体识别是信息抽取系统中最基础的工作, 是从待处理文本中找出代表现实世界中具体或抽象的实体 (人名、地名、组织机构名……) 的词语。关系抽取则是在此基础上识别出命名实体间存在的语义关系, 比如将“联想集团总裁杨元庆”语句中“联想集团”和“杨元庆”2 个实体间存在的雇佣关系抽取出来。这 2 项任务是完成其他任务的前提, 也是对数据进行结构化存储最重要的步骤。

由于在现实世界中, 关系的种类远比实体复杂, 在自然语句中出现的形式也更为灵活, 因此文本中实体间关系的抽取比实体识别更为困难。如何快速

有效地提取出大规模文本数据中实体间的关系并将其正确归类, 是当前信息抽取领域研究的热点。具体到常用的抽取方法上, 有学者<sup>[1]</sup>将其归为 5 类: 基于模式匹配的、基于词典驱动、基于机器学习、基于本体和混合方法。模式匹配和词典驱动依靠人工制定规则, 不仅费时费力, 而且领域可移植性极差。本体方法本身的构造极为复杂, 并且理论尚不成熟。基于机器学习的方法采用自然语言处理技术中统计语言模型为基础, 研究思路非常明确, 方法相对简单并取得不错的性能, 成为当下关系抽取的主流方法。基于机器学习的关系抽取方法按照有无标注好的训练语料可分为有监督、无监督和弱监督方法。训练语料是严格按照制定规则和格式将实体及其类别以及语句中实体间的关系和关系类别标注出来的一定规模的数据, 如 ACE 语料<sup>[2]</sup>。笔者结合该领域当前的进展对各类方法的原理、思想、算法局限性等方面进行了分析和总结, 并从 2 个方面对关系抽取的发展趋势进行展望。

### 1 有监督的关系抽取

有监督方法从训练数据集中学习模型, 对测试数据的关系类型进行预测。系统的输入空间是自然语句, 输出空间是预先定义好的关系种类的集合。

收稿日期: 2013-07-07; 修回日期: 2013-08-22

基金项目: 国家高技术研究发展计划 (863 计划) 主题项目 (2011AA120300); 湖南省自然科学基金资助项目 (11JJ4028)

作者简介: 刘方驰 (1989—), 男, 山东人, 硕士, 从事信息处理与信息系统技术研究。

因为在关系抽取任务中的载体都是无结构的自然语言,要使得机器能够识别以进行学习和预测,必须要把文本的各级语言单位进行形式化的表达。根据对语句的处理方式不同可以分为基于特征向量方法和基于核函数方法 2 类。

### 1.1 基于特征向量的方法

通过使用向量空间模型,基于特征向量方法将自然语句映射成特征向量。向量空间模型(vector space model)是一种应用广泛的将文本特征转化为数字特征的模型,该模型由 G.Salton<sup>[3]</sup>在 20 世纪 60 年代提出,现已成为自然语言处理领域中常用模型。

使用向量空间模型进行关系抽取时,将训练集和测试集中语句按照预先设定好的特征项赋予特定的特征值,形成多维的特征向量,之后根据选定的分类器训练集中的特征向量,用得到的训练模型对测试集进行预测。该方法将相似度高的语言单位归结为具有同一种语义关系,可以看成是一个分类问题。和一般的分类问题一样,基于特征向量的关系抽取需要解决 3 个基本问题:特征选择、特征权重计算和分类器选择。

特征项是向量模型的骨架,特征项的选择要求既要包含较多的语义信息来更完全地表征自然语句,又要对语义关系具有较高的区分度。特征的选取范围非常广泛,比如词法、实体、句法、语义及结构化信息等,但特征并不是越多越好。很多特征对于提高特征向量的语义含量没有帮助,却降低了其统计质量,造成向量稀疏问题和“维数灾难”<sup>[4]</sup>,使机器学习算法难以提取分类的统计特征并且降低了训练效率,因此如何进行特征选择至关重要。

特征选择是通过设定的准则滤除掉对分类贡献较小的特征。关系抽取中常用的特征选择方法有:基于统计信息的文档频率(DF)的特征提取法、 $\chi^2$  统计量法及基于信息论的信息增益法和互信息法。

通过特征选择选取的众多特征对于分类的贡献率不可能完全相同,可通过计算特征权重来衡量某个特征的重要程度。比如语句中实体的类别特征往往比实体词的长度特征更重要,有必要对两者赋予不同的权值。权重计算方法有 2 种思路:一是利用公式计算各特征的权值并排序,如布尔权重、特征频度等<sup>[5]</sup>;二是利用优化算法来搜索最优特征权向量<sup>[6]</sup>。权重的计算方法越复杂其对于特征表示的精确度越高,但计算成本也剧增。

特征选择以及权重计算方法解决了构造特征向

量的问题,下一个过程就是选择合适的分类器将具有相似特征的向量归为一类。具有代表性的应用于关系抽取的分类器有 K 近邻(KNN)分类器、支持向量机(SVM)、最大熵(MaxEnt)等。

Kambhatla<sup>[7]</sup>使用最大熵模型将词法、语法以及语义特征组合起来抽取语义关系。Zhao 与 Grishman<sup>[8]</sup>将文字特征,语句解析和深层语法依存特征组合,并使用 SVM 进行分类。Zhou<sup>[9]</sup>在使用句法解析树特征的同时将最基本的组块信息特征组合,来提高关系抽取的效能。Sun<sup>[10]</sup>在常用的上下文特征的基础上加入了 2 个实体间的长期相关性特征,实体顺序特征,实体间顺序特征以及标点符号特征,采用贝叶斯和感知器的混合算法进行分类。Jiang<sup>[11]</sup>系统地选取词法、句法等各种类型的特征,并且评估了不同特征的分类贡献,他们发现每一小类特征中选取最基本的单位特征就可以达到相当高的精度,而不同特征间如果不相互独立的话,反而会影响最后的性能。

另外有些研究者也用到了语句序列的各种特征,但并没有使用它们构造特征向量,而是采用其他统计方法实现关系抽取。Miao 等人<sup>[12]</sup>通过引入字特征,分词特征,句法特征,将关系抽取转化为标注问题,采用 CRF(条件随机场)方法进行估计。

基于特征向量方法是目前关系抽取最常见的方法,通过选择适当的特征可以获得很好的性能,但该方法存在着自身的局限性。首先特征向量方法致力于寻找有效的特征,对于语法结构的应用十分有限<sup>[13]</sup>。Cristianini.N 等人<sup>[14]</sup>通过研究表明,用有限维的特征向量来表示某些特定语句的句法信息是不可能的。其次,特征向量方法提高性能的方式是通过不同特征的组合,但当前使用的特征已经覆盖了大多数可以利用的语言现象,性能的提升较为困难。

### 1.2 基于核函数的方法

针对特征向量方法的局限性,部分学者使用核函数来避免上述缺陷。它不需要构造特征向量,而是直接使用字符串的原始形式作为处理对象,来计算任何 2 个对象间的核相似度(Kernel similarity)函数。基于核函数的方法可使实体关系的表示更加灵活,通过核函数的映射还可综合多方面的知识信息。

核函数是指一个函数  $K$ ,使得所有定义域里的  $x, z \in X$ ,  $X$  是输入空间

$$k(x, z) = \langle \phi(x), \phi(z) \rangle$$

这里的  $\phi$  是从空间  $X$  到内积特征空间  $F$  的映射。

核函数将原始空间中的数据点映射到一个新的特征空间，在该特征空间中训练线性分类器，而且潜在地避开了具体的计算特征映射的过程<sup>[15]</sup>，其本质是将句子背后隐式的特征向量投影到特征空间，通过计算投影的内积来表示输入空间特征向量的相似性，进而判断实体间关系的相似性<sup>[16]</sup>。

核函数的类型很多，比如多项式核函数，向量空间核函数，P-光谱核函数，全序列核函数等，并且核函数对于线性变换是封闭集合，因此利于该性质可以将多个不同信息来源的个体核函数进行复合来设计出适合特定任务的核函数，这也是该方法灵活性的一个重要体现。

使用核函数方法来抽取实体关系一般分为3个过程<sup>[16]</sup>：

1) 选择合适的解析结构来承载语句中隐含的特征信息，如使用语法树等剖析语句；

2) 在解析结构基础上选择适合的基础核函数，定量地计算解析结构中子成分的相似性；

3) 为了充分利用各种特征，可以将多个核函数复合，提高分类精度。

2003年，Zelenko.D<sup>[17]</sup>首次使用核函数方法进行关系抽取，他首先使用自然语言处理工具对语句进行标注和浅层语法分析，在此基础上构建树核函数，通过使用连续子树核和稀疏子树核函数递归的计算2个子树的相似度，来提取个人从属关系和机构位置关系，并且获得了较高的准确率。

2004年，Culotta等<sup>[18]</sup>在Zelenko提出的树核函数的基础上使用了依存树核函数，并引入知识库WordNet，扩展子树节点间的匹配函数。

2005年，Bunescu等<sup>[19]</sup>使用最短路径依存核函数来计算2个命名实体间的依存路径，实验证明最短依存路径信息适合于进行关系分类，但它要求2个实体对之间的最短依赖路径相同，因此限制了它的应用。次年，2人提出使用子序列核函数来抽取关系<sup>[20]</sup>，他们没有使用复杂的语法结构，而仅将最基本的词法信息引入核函数，不需要大量的预处理过程，并具有极快的训练速度，但性能还有待提高。

2006年，Zhang等人<sup>[21]</sup>首次使用复合核函数方法进行关系抽取，他们将卷积树核函数(convolution parse tree kernel)语实体特征核函数复合组建成一个新函数，实验表明复合核函数的表现比任一单一核函数要好。此后学者的研究主要2个方面：一是不断完善现有核函数或者发现新的核函数，如Nguyen<sup>[22]</sup>和Reichartz<sup>[23]</sup>；二是利用现有核函数组

建出性能更好的复合核函数，如Choi.S等<sup>[24]</sup>。

在中文关系抽取领域，刘克彬等<sup>[25]</sup>通过在词序列核里加入语义信息组成语义序列核函数，结合KNN机器学习算法构造分类器来分类并标注关系的类型。张骥等<sup>[26]</sup>将表征实体相关特征相似度的实体核函数和表征上下文文本相关相似度的字符串语义相似度核函数复合，设计出更能全面体现2个关系实例相似度的而复合核函数，并且设计出了基于实体形态树的核函数。

核函数可以充分搜索长距离特征和结构化特征<sup>[27]</sup>，比特征向量方法更能够充分利用语义信息。但是利用文本的解析树仍然不能彻底的表达词性、语义语境等方面的含义，不断开发新的解析结构是核方法的一个努力方向。另外，核方法将不同核函数复合可以表达高维甚至无穷维的特征空间<sup>[28]</sup>，但这也造成基于核函数方法学习和训练速度都比较慢，不适合大规模的数据抽取，同时核函数的选择多基于经验判断，缺乏严谨的理论支撑。

### 1.3 有监督方法总结

由于有监督方法以标注完好的训练集为依据，可以在同类型的测试集上取得很好的准确率和召回率。但与此同时，有监督方法又受制于训练语料，无法识别出训练语料中不包含的语义关系，因此该方法并不适合处理开放域的大规模语料。有监督方法相对成熟，其充分利用语句中各种特征信息的方式方法值得无监督等方法借鉴。

具体到2类有监督方法的对比，如果说基于特征的方法是最大限度的表征单个学习实例，那么基于核函数的方法则是尽可能地挖掘出任意2个学习实例的相似性。前者的出发点是不断寻找具有区分度的特征，后者则为选择和创造使得同类关系2个实例的相似度更大的核函数而努力。前者也使用核函数，比如在SVM中用到的线性核，多项式核等，但是在向量层面上使用核函数，后者核函数的选择范围更加宽泛，可以再字符串层面进行处理。

## 2 无监督的关系抽取

无监督方法不需要任何预先处理的语料支撑，能自动地提取文本中包含的实体关系。无监督方法可以克服人工设定关系模式的弊端，避免遗漏设定的关系模式中不包含的关系实例，而且由于不依赖于特定的训练语料，该方法对各领域的适应性很高。由于无监督方法是先确定关系实体对再确定关系类

型,因此也叫做不定义关系类别的关系抽取方法。

无监督方法包含实体对聚类 and 关系标记 2 个过程,即首先采用某种聚类方法将语义相似度高的实体对聚为一类,再选择具有代表性的词语来标记这类关系。

无监督的关系抽取方法最早是由 Hasegawa<sup>[29]</sup>在 2004 年的 ACL 会议上提出的,之后的方法多是在 Hasegawa 的基础上改进而来的。Hasegawa 方法大致分为以下步骤:首先,提取出语句中的命名实体对,把其中出现频率很低的实例剔除,将保留下来的同一命名实体对所有实例的上下文合并作为该实体对的上下文;然后,使用全联通聚类(complete linkage)方法对实体对的上下文进行聚类;最后,在同一类中寻找出现频率最高的词语作为该类命名实体对的关系描述。

Hasegawa 引出了无监督关系抽取方法都遵循的假设:即有着相同语义关系的实体对的上下文语境也具有相似性,上下文集合代表该实体对的语义关系。基于这种假设,无监督关系抽取可分为 3 个过程<sup>[30]</sup>:

- 1) 命名实体及其上下文的获取;
- 2) 对命名实体对进行聚类;
- 3) 选择核心词汇标注各类的语义关系。

但是该假设并不能保证完全正确,因为最常出现的那些实体对中可能包含一些实例,它们具有多种关系。针对这一问题,Binjamin 等<sup>[31]</sup>提出了 2 种解决办法:一种是限定进行聚类的实体对来自同一领域的语料库;另一种是通过实体的统计特性过滤掉具有多种关系的候选实体对,比如,如果 2 个实体单独出现的概率较小而同时出现的概率很大,这种实体对往往只有一种关系。同时 Binjamin 用更为灵活的基于上下文特征的模式(patterns)来代替实例的上下文,极大地提高了抽取的准确率。

张志田<sup>[32]</sup>对 Hasegawa 的方法进行了改进,把实体对的上下文特征词利用一定方法加权形成特征向量,并使用改进的 K 均值算法进行关系聚类,在 ACE 语料上测试发现性能优于 Hasegawa 方法。Edgar Gonzalez<sup>[33]</sup>将判断实例相似性的过程改进为判断是否相关,提出一种面向大规模语料的无监督关系抽取方法。Shinyama 和 Sekine<sup>[34]</sup>提出一种多层次聚类的无监督关系抽取方法。首先经过网络爬虫获得新闻文本,并且按文章的来源进行初始的分类。之后,根据语句的语义结构图,在满足一系列约束的情况下抽取出基础模式聚类的实体,这些实体根

据基本模式进行映射,形成次生聚类,如此循环往复,每个次生聚类里都包含具有相同关系的实体对。

总体而言,无监督方法不需要人工的干预指导,具有良好的领域移植性,在处理大规模开放领域数据时具有其他方法无法比拟的优势。但当前无监督方法判别实例的相似性仅用到实体对的上下文特征,对其他包含语义信息的特征未加利用,同时聚类方法的选择具有盲目性,致使无监督方法的准确率和召回率都比较低。

### 3 弱监督的关系抽取

有的监督方法具有较高的准确性,但是过分依赖标注语料,不能够有效地提取语料中不包含的关系实例,无监督方法尽管领域可移植性强但是现阶段性能较差而且无法确定提取出别的关系类别。在这种情况下,人们提出了弱监督方法的概念,即使用少量的标注语料作为初始种子集,并充分利用大规模的未标注语料来获取较好性能。具体来说,弱监督方法以预先定义关系模式和关系实例作为种子,通过机器学习,发现新的关系模板和实例。

弱监督也称为远监督(distant supervision)<sup>[35]</sup>,用这种方法提取实体关系大致包括 2 个过程:寻找包含已知关系的 2 个实体的语句,以这些语句作为训练集进行分类。

最早的弱监督方法是 Brin<sup>[36]</sup>于 1998 年提出的 DIPPRE 方法,利用实体间的关系(Relation)和描述这些关系的模式(Pattern)间的对应关系,从一个种子关系集合出发,从 Web 网页中发现这些种子出现的上下文,然后从这些上下文中产生对应的模式,利用这些模式从 Web 网页中发现更多的关系实例,然后,从这些关系实例中选择新的种子集合,重复上述过程,迭代地从 Web 上得到相应关系和模式。

Agichtein 等<sup>[37]</sup>在 Brin 基础上提出了 Snowball 系统。Snowball 在关系抽取时标注了命名实体的类型,通过限定 2 个命名实体类型的方法来改进 Brin 的方法,并且 Snowball 对模式的定义更加灵活。但是 Snowball 仅针对于单一的关系类型,2010 年,Liu<sup>[38]</sup>提出可以提取多类型关系的架构:MultiSnowball,该系统致力于从一个提取模式中抽取不同的关系类型并且可以迭代地发现新的关系类型,同时可以使不同关系类型间共享模式。

以上方法都是基于自举(Bootstrapping)的思想,即通过现有模式不断扩展出新的模式,属于启发式的方法,文献[39-41]对启发式方法做了改进,

但这类方法的缺陷是在扩展过程中很多不正确的模式进入模式集中，并且使得错误不断累积，影响后面的关系分类。学者试图通过研究新的方法来避免上述缺陷。

2006年，Jinxiu Chen等<sup>[42]</sup>在ACL会议上提出标注传播算法(Label Propagation)，这是一种基于图的弱监督学习方法。将所有的实体对看作是图上的节点，将实体对间的距离看作边。把一部分标注好的节点看作源头向其他节点传播，而权重值越高的边上传播的速度越快。将相似度高的节点聚为一类，类别则息通过传播过来的标注信息来判别。

2011年，A.Cvitas<sup>[43]</sup>提出了一种新的弱监督方法——协同学习(Co-learning)方法，其基本流程：选择2个不同的分类器，使用相互独立的特征在2个训练集上训练，并分别在未标注集上测试，选取置信度高的实例扩展到另一个分类器的训练集中，如此迭代若干次，当精度达到阈值时停止。协同学习方法的关键是寻找2个或多个相互独立的特征集，M.F.Moens等<sup>[44]</sup>在这方面进行过研究，核心思

想是根据特征相对于2个实体的位置分类，这为协同学习方法引入多个分类器进行扩展提供了依据。

弱监督方法结合了前2种方法的优势，不需要过多的训练语料，并且可以发现种子集中不包含的关系实例，因此成为关系抽取研究的热点。很多学者将在其他领域表现不俗的算法或思想引入到弱监督方法中，使其更加完善，如Enrique<sup>[35]</sup>引入主题模型(Topic Model)来计算模式隶属于某类关系的概率，Mike Mintz<sup>[45]</sup>尝试使用知识库FreeBase来代替手工选取种子集等。但弱监督方法的2个主要问题<sup>[46]</sup>：如何减少选取种子集的人工干预以及如何控制扩展过程中的噪声并没有完全解决，这也限制了该方法的性能。

笔者将基于机器学习的实体关系抽取3类方法的相关内容和研究进展进行了分析，并指出各种方法的优势与不足。实体关系抽取是信息抽取中的关键任务，根据数据资源和任务需求合理选择抽取方法是实现关系抽取的核心所在。表1是对3类方法几个方面的比较总结。

表1 3类方法对比

项目	领域移植性	人工干预	研究思路	性能	性能提升方法
有监督	弱	强	分类	高	改进特征和核函数
弱监督	中	弱	分类	中	改进模式扩展方法
无监督	强	弱	聚类	低	扩展特征，改进聚类算法

## 4 结束语

现今的关系抽取方法多面向于1个语句中的2个实体，对于跨句子、段落甚至跨篇章的2个或多个实体间复杂关系的研究很少，这也是阻碍关系抽取达不到真正实用的一个原因。其次，现有的实体关系类型都是人工制定的，如何使系统自动发现关系类型是今后关系抽取研究的一个方向。另外，有监督方法已经比较成熟，现有的方法多是对经典算法的改进和补充，难有太大的突破，而且其基于训练语料的思想使它面向开放领域时的领域迁移能力较弱。随着数据和各类信息的爆炸式增长，面向特定类型或特定领域的关系抽取已经难以满足要求，无监督和弱监督方法必然会成为主流。不断提高这2种方法的准确率和召回率是今后研究的重点。

## 参考文献：

- [1] 徐建, 张智雄, 吴振新. 实体关系抽取的技术方法综述[J]. 现代图书情报技术, 2008, 168(8): 18-23.
- [2] www ldc.upenn.edu/Projects/ACE.
- [3] Salton G., McGill M J. Introduction to Modern Information Retrieval [M]. New York: McGraw-Hill Book Company,

1983: 400-403.

- [4] 宗成庆. 统计自然语言处理[M]. 北京: 清华大学出版社.
- [5] 鲁松, 李晓黎, 白硕, 等. 文档中词语权重计算方法的改进[J]. 中文信息学报, 2000, 4(6): 8-14.
- [6] 赵宇. 关系抽取中学习算法的研究[D]. 哈尔滨: 哈尔滨工业大学, 2006.
- [7] Kambhatla N., Combining lexical, syntactic and semantic features with Maximum Entropy models for extracting relations[C]. In Proceedings of ACL'2004, 2004, Barcelona, Spain.
- [8] Zhao S., Grishman R. Extracting relations with integrated information using kernel methods[C]. In Proceedings of the ACL'2005, 2005, Ann Arbor, USA.
- [9] Zhou G. D., Su J., Zhang J., et al. Exploring various knowledge in relation extraction[C]. In Proceedings of the ACL'2005, 2005, Ann Arbor, USA.
- [10] Sun Xia, Dong Lehong. Feature-based Approach to Chinese Term Relation Extraction[C]. 2009 International Conference on Signal Processing Systems, 410-414.
- [11] Jiang Jing, Zhai Chengxiang. A Systematic Exploration of the Feature Space for Relation Extraction[C]. In proceedings of NAACL/HLT, 2007: 113-120.
- [12] Miao Qingliang, Zhang Shu, Zhang Bo, et al. Extracting and Visualizing Semantic Relationships from Chinese Biomedical Text[C]. 26th Pacific Asia Conference on Language, Information and Computation, 99-107.
- [13] Jung Hanmin, Sung-Pil Choi, Seungwoo Lee, et al.

- Survey on Kernel-Based Relation Extraction[S.].
- [14] Cristianini. N, Shawe Taylor. J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods[C]. Cambridge University Press.
- [15] Scholkopf, Smola A. J. Learning with Kernels: SVM, Regularization, Optimization and Beyond[C]. MIT Press, Cambridge, MA, 2001: 407-423.
- [16] Zhang Xiaofeng, Gao Zhiqiang, Man Zhu. Kernel methods and its application in Relation Extraction. 2011 IEEE, 1362-1365.
- [17] Zelenco. D, Aone. C, Richardella. A. Kernel Methods for Relation Extraction[J]. Journal of Machine Learning Research, 3, 1083-1106.
- [18] Culotta. A, Sorensen. J. Dependency Tree Kernels for Relation Extraction [C]. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics.
- [19] Bunescu. R., Mooney. R. J. A Shortest Path Dependency Kernel for Relation Extraction[C]. Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, 724-731.
- [20] Bunescu. R, Mooney. R. J. Subsequence Kernels for Relation Extraction[C]. Proceeding of the Ninth Conference on Natural Language Learning, 2005.
- [21] Zhang. M, Zhang. J, Su. J, et al. A Composite Kernel to Extract Relations between Entities with both Flat and Structured Features[C]. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL, 2006: 825-832.
- [22] Nguyen. T, Moschitti. A, Riccardi. G. Convolution kernels on constituent, dependency and sequential structures for relation extraction[C]. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 2009, 1378-1387.
- [23] Reichartz. F, Korte. H, Paass. G. Dependency tree kernels for relation extraction from natural language text[J]. Machine Learning and Knowledge Discovery in Databases, 2009: 270-285.
- [24] Choi. S, Jeong. C, Choi. Y, et al. Relation Extraction based on Extended Composite Kernel using Flat Lexical Features[J]. Journal of KIISE: Software and Applications, 2009, 36(8).
- [25] 刘克彬, 李芳, 刘磊, 等. 基于核函数中文关系自动抽取系统的实现[J]. 计算机研究与发展, 2007, 44(8): 1406-1411.
- [26] 张骥. 基于核函数的中文实体关系抽取新方法[D]. 天津: 天津大学.
- [27] 林如琦, 陈锦绣, 杨肖方, 等. 多信息融合中文关系抽取技术研究[J]. 厦门大学学报, 2011, 5(3): 540-545.
- [28] Shawe Taylor J., Cristianini N. Kernel Methods for Pattern Analysis[C]. Cambridge University Press, Cambridge, 2004.
- [29] Hasegawa. T, Sekine. S, Grishman. R. Discovering Relations among Named Entities from Large Corpora[C]. In Proc. of ACL-2004, 2004: 415-422.
- [30] 邸楠. Web 中实体关系分析与面向实体的社会网络分析研究[M]. 北京: 北京大学网络实验室.
- [31] Benjamin Rozenfeld, Ronen Feldman. High-Performance Unsupervised Relation Extraction from Large Corpora[C]. Proceedings of the Sixth International Conference on Data Mining (ICDM'06), 2006.
- [32] 张志田. 无监督关系抽取方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2007.
- [33] Edgar Gonzalez, Jordi Turmo. Unsupervised Relation Extraction by Massive Clustering[C]. 2009 Ninth IEEE International Conference on Data Mining, 782-787.
- [34] Shinyama Y., Sekine S. Preemptive information extraction using unrestricted relation discovery[C]. in Human Language Technology Conference and North American chapter of the Association of Computational Linguistics Annual Meeting (HLT-NAACL), 2006: 304-311.
- [35] Enrique Alfonseca, Katja Filippova, Jean-Yves, et al. Pattern Learning for Relation Extraction with a Hierarchical Topic Model[C]. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 54-59.
- [36] Brin S. Extracting patterns and relations from the world wide web[C]. WebDB Workshop at 6th International Conference on Extending Database Technology, EDBT98, 1998: 172-183.
- [37] Agichtein E., Gravano L. Snowball: extracting relations from large plain-text collections[C]. Proceedings of the fifth ACM conference on Digital libraries, 2000: 85-94.
- [38] Liu Xiaojiang, Yu Nenghai. MultiType Web Relation Extraction Based on Bootstrapping[C]. 2010 WASE International Conference on Information Engineering, 2010: 24-27.
- [39] Mintz M., Bills S., Snow R., et al. Distant supervision for relation extraction without labeled data[C]. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2:1003-1011.
- [40] Hoffmann. R, Zhang. C, Ling. X, et al. Knowledge-based weak supervision for information extraction of overlapping relations[C]. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011: 541-550.
- [41] Wang C., Fan J., Kalyanpur A., et al. Relation extraction with relation topics[C]. In Proceedings of Empirical Methods in Natural Language Processing, 2011.
- [42] Chen Jinxiu, Dong Hong, et al. Relation Extraction Using Label Propagation Based Semi-supervised learning[C]. Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL.
- [43] Cvitas A. Relation Extraction from Text Document[J]. MIPRO 2011, Opatija, Croatia, 2011: 23-27.
- [44] Moens M. F. Information Extraction: Algorithms and Prospects in a Retrieval Context[R]. Springer, 2006
- [45] Mike Mintz, Steven Bills, Rion Snow, et al. Distant supervision for relation extraction without labeled data[C]. Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, 1003-1011.
- [46] 毛小丽. 基于特征向量的实体间语义关系抽取研究[D]. 重庆: 重庆大学, 2011.