

# 关系抽取综述

谢德鹏<sup>1</sup>, 常青<sup>2</sup>

(1. 中国航天科工集团第二研究院研究生院, 北京 100854; 2. 北京计算机技术及应用研究所, 北京 100854)

**摘要:** 关系抽取任务作为信息抽取的基本组成之一, 在很多领域具有十分重要的地位。关系抽取发展至今, 总体可以分为基于规则的抽取方式和基于统计方式的抽取; 之后出现的众多方法大多是以统计为主, 辅助以规则。后来引入了包括远程监督、深度学习等模式并融合了注意力机制、多标签多实例方法。接下来将会对关系抽取的发展过程和方向及以上提到的方法进行介绍。

**关键词:** 关系抽取; 有监督方法; 无监督方法; 半监督方法; 远程监督; 神经网络; 联合抽取

**中图分类号:** TP391.4      **doi:** 10.19734/j.issn.1001-3695.2018.12.0923

## Review of relation extraction

Xie Depeng<sup>1</sup>, Chang Qing<sup>2</sup>

(1. Graduate School, Second Research Academy of CASIC, Beijing 100854, China; 2. Beijing Institute of Computer Technology and Application, Beijing 100854, China)

**Abstract:** As one of the basic part of information extraction, relation extraction plays an important role in many fields. And approaches generally consist of rule-based and statistic-based. Many subsequent methods are mainly based on statistics and supplemented by rules. Some methods are applied including distant supervision, deep learning and etc. Meanwhile, methods are integrated including attention mechanism, multi-tag and multi-instance. Next, this paper reviewed the history and the methods mentioned above, and indicated some promising directions.

**Key words:** relation extraction; supervised relation methods; unsupervised relation methods; semi-supervised relation; distant supervision; neural network; joint extraction

## 0 引言

伴随着互联网技术的不断发展, 网络资源也呈现出爆炸式地飞速增长, 出现更多的是半结构化和非结构化的数据资源。如何低成本地从复杂、多源分散和冗余异构的数据中提取出高质量的、结构化的可利用数据成为现实和迫切的问题, 这也是信息抽取的主要研究内容。关系抽取是信息抽取的子任务之一。

关系抽取包括实体及其关系抽取两部分, 主要是从文本中抽取实体 1 和实体 2 及其两者之间的关系类型, 表示为类似于〈实体 1, 关系类型, 实体 2〉的结构化形式。现在主流的抽取方法可以细分为基于监督的方式、基于半监督的方式、无监督方式、远程监督和面向开放域的关系抽取, 并且深度学习加入之后逐渐替代传统机器学习方法成为研究的热点和主流。

关系抽取在诸多领域都具有重要的应用价值, 包括语义理解、推荐检索、机器翻译和智能问答等服务, 关系抽取都作为基础数据提供的手段。伴随着知识图谱研究和应用的盛行, 关系抽取作为对图谱构建最终的生成质量具有举足轻重影响的基本和核心环节也是备受关注, 吸引大批的学者和研究者参与其中, 很多引入和融合的新方法和模型也就此产生。

## 1 关系抽取的产生发展

最早在 1998 年的第七届信息理解会议 (message understanding conference, MUC) 上, 首次提出了模板抽取的任务, 这在后来发展成为关系抽取。在 MUC-7 会议上提出的关系抽取任务主要针对人物 (persons)、地理位置

(locations) 和产品 (artifacts) 的模板类型<sup>[1]</sup>, 评测的语料内容主要来源于纽约时报对飞机事故和航天发射相关的新闻报道, 并且设计了抽取结果的评价体系。

在 1999 年, 美国国家技术研究院(National Institute of Standards and Technology, NIST)召开的自动内容抽取会议 (automatic content extraction, ACE)取代了之前的 MUC 会议, 并且对关系抽取评测的任务和训练材料进行了融合拓展和细化完善<sup>[2]</sup>。在 2002 年, 第三届会议正式加入了实体关系发现和识别任务(relation detection and recognition, RDR)。2008 年, ACE 会议正式将关系抽取任务划分为七种类型。2009 年, ACE 正式纳入文本分析会议(text analysis conference, TAC), 成为知识库总体 (knowledge base population, KBP) 任务的重要组成部分<sup>[3]</sup>。后期出现的语义评估 (semantic evaluation, SemEval) 会议 SemEval-2007 的评测任务 4 中设置了七种常用名词和名词短语间的实体关系, 在 SemEval-2010 评测任务 8 中将实体关系类型扩充到了 10 种<sup>[4]</sup>。

以上提到的包含关系抽取研究的会议极大地推动了其发展, 但是它们所发布的评测语料对于人工标注的依赖性较大。这类语料库耗费大量人力进行手工的模板和规则的编写及训练语料的标注, 虽然质量有所保证, 却无法提供大规模材料, 并且领域适应性和后期扩充性很差。后来包括维基百科、DBpedia 和 Freebase 等大规模事实知识库出现后, 为标注语料提供大量的语料支持, 使得开放域关系抽取成为可能, 并使其具有在跨领域性和规模性方面的先天优势。

## 2 关系抽取的研究现状

关系抽取作为信息抽取中的重要子任务, 国内前期的相

收稿日期: 2018-12-28; 修回日期: 2019-02-27

作者简介: 谢德鹏 (1993-), 男, 山东临沂人, 硕士研究生, 主要研究方向为关系抽取、知识图谱(1181780298@qq.com); 常青 (1969-), 女, 天津人, 副总工程师, 研究员, 硕导, 硕士, 主要研究方向为计算机应用。

关研究起步较晚, 后期包括中国科学院、清华大学、北京大学等诸多科研院所在推动其研究发展上作出重要工作。在实现方式上, 根据对于人工标注数据的依赖性可以细分为基于监督的方式、基于半监督的方式、无监督方式和面向开放域的抽取。后期又出现远程监督方式, 深度学习逐渐火热和成熟之后也被应用到关系抽取之中, 取得相比传统机器学习方法更加优秀的效果。接下来将分别对这些方法的研究进展进行详细的介绍。

## 2.1 有监督的关系抽取

基于监督的关系抽取以较高的准确率成为目前业界广泛应用的方式。其主要是利用分类的思想根据已有的人工标注数据进行模型的训练, 然后进行特定关系的匹配识别和抽取工作。监督方式主要分为两大方法体系, 分别是基于特征向量 (feature-based) 的方法和基于核函数 (kernel-based) 的方法。

基于特征向量的方法主要通过从句子上下文中提取出包括句法和语法等特征信息去构造特征向量, 进而利用特征向量的相似度训练实体关系识别模型, 完成实体关系识别和抽取。Kambhatla<sup>[5]</sup>利用最大熵分类器构建抽取模型, 通过加入文本特征, 可以使用很少的词汇特征达到不错的效果, 从而降低对于语义特征提取树的依赖, 模型在 ACE RDC 2003 英文语料上的关系抽取的 F 值为 52.8%。很多基于特征向量的方法借助于传统机器学习实现, 并且对特征的选取有很大增益。Giuliano 等人<sup>[6]</sup>基于 SemEval-2007 的评测材料, 通过实体上下文、距离等特征, 借助支持向量机 (support vector machine, SVM) 模型实现 71.8% F 值的抽取效果。Tratz 等人<sup>[7]</sup>利用最大熵分类方法在 SemEval-2010 评测语料上实现抽取效果 F 值为 77.57%。Culotta 等人<sup>[8]</sup>使用条件随机场的分类方法借助带有默认正则化参数的 MALLET CRF 实现关系抽取, 取得了不错的效果。

虽然基于特征向量的抽取方法效果良好, 但是作为模型基础的特征在选择和设置上更多依靠构建者的直觉和经验, 并且对于上下文信息利用不足, 于是需要提出可以较好地利用语料中的长距离特征和结构化特征的函数方法, 在这一层面进行了弥补。Zelenko 等人<sup>[9,10]</sup>引入基于核函数的关系抽取方法, 通过核函数实现从低维向高维空间的映射, 从而可以将非线性问题作为线性问题处理。Zelenko 等人<sup>[9]</sup>提出使用浅层解析树核结合支持向量机从文本中提取人一隶属关系和组织一位置关系。Culotta 等人<sup>[11]</sup>使用依存树核对新闻文章自动内容抽取 (ACE) 语料库中实体之间的关系进行了检测和分类, 并且测试了词性和实体类型不同特性的效用, 证实依赖树内核比“词袋”核实现了 20% F1 的改进。Bunescu 等人<sup>[12]</sup>观察到依赖关系图中两个实体之间的最短路径可查找到实体关系, 所以利用最短依存树核进行改进。Zhang 等人<sup>[13]</sup>提出基于解析树的卷积核对句法结构信息建模, 实现关系抽取的方法。庄成龙等人<sup>[14]</sup>通过在原关系实例的结构化信息中增添实体语义信息并筛除冗余信息的方法来提高关系抽取的性能, 在 ACE RDC 2004 基准语料上进行的关系检测, F 值达到了 79.1%。总体上, 基于核函数的方法由于匹配计算时的较强约束容易出现召回率较低的现象, 并且模型训练和预测的时间复杂度较高。

## 2.2 半监督的关系抽取

半监督的关系抽取方法的主要思想是根据预先设计好的关系类型, 通过人工添加合适的实体对作为种子。利用模式学习方法进行不断迭代学习, 最终生成关系数据集和序列模式, 在一定程度上降低了对于人工标注语料的依赖。

最常用途径是基于 Bootstrapping 方法实现。这是由 Brin<sup>[15]</sup>首先应用在关系抽取任务中, 并建立了 DIPRE 系统, 他

以少量的书名及作者名作为种子实体关系对, 从文档和语句中抽取新的实例并作为标注样本, 根据标注样本建立新的抽取模板。利用建立的模板发现新的实体对关系并加入其中, 期间不断调整和迭代。Agichtein 等人<sup>[16]</sup>设计的 Snowball 方法是在其基础上推出的, 主要改进是使用向量表示实体及实体关系的元组, 通过向量的相似度来发现和迭代增大标注样本。在每次迭代提取的过程中, 模型在不受人工干预的情况下评估这些模式和元组的质量, 并且只保留其中最可靠的部分以提升整体质量。陈锦秀等人<sup>[17]</sup>利用图策略建立基于图的半监督抽取模型, 实现关系抽取性能的提升。

半监督的方法降低了关系抽取对于人工标注数据的依赖性, 只需要人工进行最初种子集的构造, 但是对种子集的质量要求较高, 并且建立和优化相对繁琐的模板对于最终的抽取效果至关重要。这种方式普遍存在噪声实例及模板引入问题, 进而在不断迭代过程中造成语义漂移的现象; 且此类方法虽准确率有所提高, 但是召回率普遍不高。

## 2.3 无监督的关系抽取

无监督的方法属于一种自底向上的抽取方式, 通过先抽取实体及关系, 进行大规模的冗余语料聚类, 再对聚类集合进行关系标注。Hasegawa 等人<sup>[18]</sup>首次应用无监督方式进行关系抽取, 通过设置重复出现阈值识别潜在语义关系并聚类, 实现抽取结果 F 值达到 75%。Shinyama 等人<sup>[19]</sup>基于多层次聚类的无监督方法, 使用了 12 家主要在美国出版的报道文章进行实验。

Hassan 等人<sup>[20]</sup>提出了一种基于大数据集冗余和图相互增强的无监督信息提取方法, 并采用从语料库中的 POS n-gram 获取关系提取模式。但是由于采用 n-gram, 这种模式只包含 POS 和实体类型标记, 导致 n-gram 数量的组合激增。Gonzalez 等人<sup>[21]</sup>提出一种新的基于概率聚类模型的无监督方式关系提取方法, 该方法得到的 F1 值为 55.7。Rozenfeld 等人<sup>[22]</sup>建立的 URIES 是一个 Web 关系提取系统, 通过对目标关系及其属性的简短描述, 从未标记的文本中进行模式提取。

无监督抽取方式虽然对人工标注的语料依赖性降低, 并且多领域适应性强, 领域迁移障碍小, 特别在多领域知识杂糅的大规模文本中相比其他有监督和半监督方法更是优势明显。但是无监督方法总体上关系标注较为宽泛, 因为缺乏必要的语料库, 导致低频实例抽取率低, 最终识别的准确率和召回率也一般不高, 在抽取评价标准上也难以量化和统一。

## 2.4 面向开放域的关系抽取

面向开放域的关系抽取不限定关系类别和目标文本, 在跨领域和后期扩充上具有无法比拟的优势。开放式关系抽取默认同一实体对都存在已知的关系, 通过前后相邻的短语进行实体关系上的语义表达, 借助外部大型实体知识库包括 DBpedia、YAGO、FreeBase 和其他领域知识库, 将置信度较高的实体关系与大规模的训练数据进行匹配对齐, 以获得大量训练数据。

Etzioni 等人<sup>[23]</sup>搭建 KnowItAll 模型, 通过人工编写规则模板从 web 中进行无监督的、独立于领域和面向可伸缩的大量事实 (如科学家或政治家的姓名) 地自动化匹配。模型中每个规则由谓词、提取模式、约束和关键字组成, 进而借助简单的语法分析抽取实体关系。Banko 等人<sup>[24]</sup>通过构建 TextRunner 模型, 从 Web 中抽取包含用户输入的特定谓词的元组, 模型包括三个关键模块: 自我监督学习器对输入的小语料库样本进行处理, 输出分类器对候选提取置信度的标记; 单通道提取器利用整个语料库从每个句子生成一个或多个候选元组, 进而得到所有可能的元组, 保留标记为可信的元组; 冗余评估器依据文本中的冗余概率模型为每个保留的元组分配概率。整个过程不需要人工进行标注, 降低工作量, 但是

依旧存在召回率不高的问题。

## 2.5 应用远程监督方法的关系抽取

后来的研究着力点主要集中在如何降低对于人工标注语料库的依赖性, 增强领域迁移性上。Mintz 等人<sup>[25]</sup>在文本处理中尝试借用远程监督方法, 假设若文本中的实体对和知识库的实体对完全一致, 就标注它们具有同样的关系。该方法通过对齐语料库和文本自动生成训练样例, 从而提取特征训练分类器, 降低对于人工标注材料的依赖; 但是知识库中事先标注的实体关系是不完备的, 所以过于简易和强烈的假设极易引入错误。远程监督方法主要通过知识库与非结构化文本对齐来自动构建大量训练数据, 减少模型对人工标注数据的依赖, 增强模型跨领域适应能力。

为了改善远程监督单标签过强假设的问题, Surdeanu 等人<sup>[26]</sup>又进一步提出基于概率图模型的多标签多实例的抽取方式, 并引入词袋 (word bag) 模型, 通过提升标注级别, 将原本实体对级的标注改变为对多词形成的词袋进行标注以降低错误率。

## 2.6 基于深度学习的关系抽取

以上方法都借助传统的自然语言处理工具, 但是工具本身也是很容易引入错误, 经过这些工具处理后的结果降低了接下来的算法性能。考虑到语音、图像和文本处理肌理的相通性, 当深度学习方法在图像领域崭露头角时, 很多人尝试引入深度学习方法到关系抽取中, 发挥其在特征提取和自动学习上的优势, 并且将 SemEval-2010 task 8<sup>[27]</sup>作为测试标准。

Socher 等人<sup>[28]</sup>通过使用递归神经网络 (recurrent neural networks, RNN) 模型, 在句法树的节点上设置向量和矩阵, 对命题逻辑和自然语言中算子的含义学习, 从而得到多种句法类型和不同长度短语和句子的向量化表示。模型在 SemEval-2010 Task 8 数据库上实现 F 值为 82.2% 的抽取效果。Hashimoto 等人<sup>[29]</sup>在网络分类器中加入词嵌入方法, 从语料库中抽取实体对上下文特征信息, 基于同样的数据库实现小幅度提升。

递归神经网络关注于语义的结构信息, 但为获得这一信息需要依赖于传统自然语言处理工具, 传统自然语言处理 (natural language processing, NLP) 工具噪声引入的弊端再次显露。于是 Zeng 等人<sup>[30]</sup>利用卷积神经网络 (convolutional neural networks, CNN) 提取词汇和句子的层次特征进行关系抽取, 减少输入材料复杂的预先标记处理。Nguyen 等人<sup>[31]</sup>在此工作的基础上向卷积层中加入了多尺寸的卷积核作为过滤器, 以此提取更多的 N-Gram 特征, 并且使用了位置向量, 证实了多尺寸卷积神经网络在关系抽取中的有效性。

Lin 等人<sup>[32]</sup>引入 PCNN (piecewise CNN), 对传统卷积神经网络的池化层进行改进, 通过两个实体位置将 feature map 分为三段进行池化, 其目的是更好地捕获两个实体间的结构化信息, 并使用注意力机制, 通过建立句子级选择性注意神经模型减轻错误标签问题, 最终 F 结果比基于多示例学习的机器学习方法高了 5 个百分点。Zhou 等人<sup>[33]</sup>使用 BLSTM (bidirectional long-short term memory) 对句子建模, 并使用 word 级别的 attention 机制提升结果。万静等人<sup>[34]</sup>利用双向 GRU 和 PCNN 方法实现实体结构和更多特征信息的提取, 在 NYT 数据集上不错效果。

Cai 等人<sup>[35]</sup>提出了基于最短依赖路径 (shortest path dependence, SDP) 的深度学习关系分类模型, 称为双向递归卷积神经网络模型 (BRCNN)。使用 SemEval-2010 Task 8 数据集, 在关系分类任务中, 实现了 F 值高达 86.3% 的抽取效果。

2017 年, Lin 等人<sup>[36]</sup>又尝试通过扩展到多语言语料库上, 利用多语言语料信息的互补性和一致性提升抽取性能, 这在

多语言共存的文档中效果显著。

但是以上的所有关系抽取方法都将其分解为实体抽取 (named entity recognition, NER) 和关系抽取 (relation extraction, RE) 两个依次进行的步骤。这种分割忽略子任务间的关系, 容易产生冗余信息, 于是端到端方式的联合抽取被提出。在同一个模型中抽取出实体及其之间关系类型, 实现参数共享、同步优化, 降低之前流水式抽取出现错误累积的可能性。Zheng 等人<sup>[37]</sup>利用共享神经网络底层表达进行联合学习。Miwa 等人<sup>[38]</sup>同样通过参数共享 NER 使用一个神经网络进行解码, 在 RC 上加入了依存信息, 根据依存树最短路径使用一个 BiLSTM 来进行关系分类。Li 等人<sup>[39]</sup>提出了增量集束搜索算法的联合结构化抽取方式和利用全局特征的约束方法, 在 ACE 语料上比传统的流水线方法 F 值提升了 1.5%。Zheng 等人<sup>[40]</sup>使用了更加高效的偏置目标函数和一种新的标注策略, 把原来涉及到序列标注任务和分类任务的关系抽取完全变成了一个序列标注问题, 通过一个端对端的神经网络模型直接得到关系实体三元组。

## 3 关系抽取的评价指标

这套关系抽取工作的最终效果评价体系是在 ACE 会议上提出, 以准确率 (precision)、召回率 (recall) 进行衡量, 但是准确率和召回率在一定程度上过于偏重评测抽取的单方面效果, 于是引入 F 值, 综合衡量抽取的结果。

三者的计算公式为

$$\text{Precision} = \frac{\text{正确分类为某类的实例数}}{\text{被划分为某类的实例数}} \times 100\% \quad (1)$$

$$\text{Recall} = \frac{\text{正确分类为某类的实例数}}{\text{测试集中某类实例的总数}} \times 100\% \quad (2)$$

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (3)$$

## 4 关系抽取的挑战和趋势

总体上, 关系抽取在领域自适应性和召回率层面仍有提升空间。借助已有的知识库并挖掘深度学习语义表示和自主学习能力, 自动从训练数据中学习分类特征、自主进行语料扩充, 进而增强领域的迁移性, 减少对于人工标注语料的依赖。因此弱监督包括远程监督方式应该是研究和应用的方向之一, 但是目前远程监督的错误标注噪声引入问题一直存在, 虽通过多标签多实例方式进行改善, 但是仍有提升空间。

并且现阶段的关系抽取大多集中在词汇、语句级别的实体间关系, 很少扩大到段落甚至篇章级别的关系抽取, 但是大量的代词在文段中需要依靠上下文信息, 甚至段落篇章级进行语义理解, 指代词在语义上和名词实体的等价关系应充分利用。通过共指消解处理结果的引入, 更好地进行实体之间等价关系和非等价关系的融合, 也可以进一步推理出长文档或者多文档实体之间存在的间接和潜在关系, 借助简单推理实现长距离关系抽取, 进一步提升关系抽取的召回率。

现在关系抽取的主流方式是二元关系抽取, 即使是多元实体关系仍将其作为二元实体关系进行处理。但是现实世界中实体之间的三元甚至多元关系广泛存在, 将二元抽取方法延伸至多元抽取, 发现潜在实体关系、多元实体关系、多层次关系将会对实际应用产生巨大影响。

## 5 结束语

现阶段, 关系抽取在商业化应用上, 基于监督的方式可以抽取更加有效准确的特征信息, 以其较高的抽取准确率和召回率有着很大的优势, 但是其依赖的标注语料库构建困难、完备性差、领域适应性不足, 在新领域应用仍需要创建

抽取规则和样例, 后期扩充难度很大; 基于无监督方式虽然跨领域性强, 特别适合处理大规模、多领域复杂文本, 但是聚类的阈值等参数的设置和优化一直以来难以快速确定和实现; 半监督的方式通过少量人工标注的种子实例进行迭代地扩充训练集, 但是过程中容易引入错误标注和噪声, 进而形成语义漂移现象。基于深度学习的方法已经成为重要方向, 更好地学习和利用词性、句法等语义特征, 进一步减轻对于标注语料的依赖程度, 进而增强领域的迁移性和适应性, 因此其仍将是研究的重点方向。多层次、有偏重的注意力机制也会进一步应用以充分利用有价值信息。关系抽取作为智能检索、知识图谱、篇章标注、智能问答等研究的基础支撑技术具有十分重要的研究价值, 并且抽取效果仍有很大的提升空间。

## 参考文献:

- [1] 刘绍毓. 实体关系抽取关键技术研究 [D]. 郑州: 解放军信息工程大学, 2015. (Liu Shaoyu. Research on key technologies for entity relation extraction [D]. Zhengzhou: PLA Information Engineering University, 2015.)
- [2] 张素香. 信息抽取中关键技术研究 [D]. 北京: 北京邮电大学, 2007. (Zhang Suxiang. Research on key technologies of the information extraction [D]. Beijing: Beijing University of Posts and Telecommunications, 2007.)
- [3] Doddington G R, Mitchell A, Przybocki M A, *et al.* The automatic content extraction (ACE) program-tasks, data, and evaluation [C]// Proc of LREC. 2004.
- [4] Hendrickx I, Kim S N, Kozareva Z, *et al.* Semeval-2010 task 8: Multi-way classification of semantic relation between pairs of nominal [C]// Proc of Workshop on Semantic Evaluations: Recent Achievements and Future Directions. 2009: 94-99.
- [5] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations [C]// Proc of ACL on Interactive Poster and Demonstration Sessions. Stroudsburg, PA: Association for Computational Linguistics, 2004: 22-26.
- [6] Giuliano C, Lavelli A, Pighin D, *et al.* FBK-IRST: kernel methods for semantic relation extraction [C]// Proc of the 4th International Workshop on Semantic Evaluations. Stroudsburg, PA: Association for Computational Linguistics, 2007: 141-144.
- [7] Tratz S, Hovy E. Isi: automatic classification of relations between nominals using a maximum entropy classifier [C]// Proc of the 5th International Workshop on Semantic Evaluation. Stroudsburg, PA: Association for Computational Linguistics, 2010: 222-225.
- [8] Culotta A, McCallum A, Betz J. Integrating probabilistic extraction models and data mining to discover relations and patterns in text [C]// Proc of Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2006: 296-303.
- [9] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction [J]. Journal of Machine Learning Research, 2003, 3 (3): 1083-1106.
- [10] Zhao Shubin, Grishman R. Extracting relations with integrated information using kernel methods [C]// Proc of the 43rd Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2005: 419-426.
- [11] Culotta A, Sorensen J. Dependency tree kernels for relation extraction [C]// Proc of the 42nd Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2004: 423.
- [12] Bunescu R C, Mooney R J. A shortest path dependency kernel for relation extraction [C]// Proc of Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2005: 724-731.
- [13] Zhang Min, Zhang Jie, Su Jian. Exploring syntactic features for relation extraction using a convolution tree kernel [C]// Proc of Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2006: 288-295.
- [14] 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. (Zhuang Chenglong, Qian Longhua, Zhou Guodong. Research on tree kernel based entity semantic relation extraction [J]. Journal of Chinese Information Processing, 2009, 23 (1): 3-9.)
- [15] Brin S. Extracting patterns and relations from the World Wide Web [J]. Lecture Notes in Computer Science, 1998, 1590: 172-183.
- [16] Agichtein E, Gravano L. Snowball: extracting relations from large plain-text collections [C]// Proc of ACM Conference on Digital Libraries. New York: ACM Press, 2000: 85-94.
- [17] 陈锦秀, 姬东鸿. 基于图的半监督关系抽取 [J]. 软件学报, 2008 (11): 2843-2852. (Chen Jinxiu, Ji Donghong. Graph-based semi-supervised relation extraction [J]. Journal of Software, 2008 (11): 2843-2852.)
- [18] Hasegawa T, Sekine S, Grishman R. Discovering relations among named entities from large corpora [C]// Proc of the 42nd Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2004: 415-422.
- [19] Shinyama Y, Sekine S. Preemptive information extraction using unrestricted relation discovery [C]// Proc of Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2006: 304-311.
- [20] Hassan H, Hassan A, Emam O. Unsupervised information extraction approach using graph mutual reinforcement [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2006: 501-508.
- [21] Gonzalez E, Turmo J. Unsupervised relation extraction by massive clustering [C]// Proc of the 9th IEEE International Conference on Data Mining. Piscataway, NJ: IEEE Press, 2009: 782-787.
- [22] Rozenfeld B, Feldman R. High-performance unsupervised relation extraction from large corpora [C]// Proc of the 6th International Conference on Data Mining. Piscataway, NJ: IEEE Press, 2006: 1032-1037.
- [23] Etzioni O, Cafarella M, Downey D, *et al.* Unsupervised named-entity extraction from the Web: an experimental study [J]. Artificial intelligence, 2005, 165 (1): 91-134.
- [24] Banko M, Cafarella M J, Soderland S, *et al.* Open information extraction from the Web [C]// Proc of the 20th International Joint Conference on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann Publishers Inc, 2007: 2670-2676.
- [25] Mintz M, Bills S, Snow R, *et al.* Distant supervision for relation extraction without labeled data [C]// Proc of the 47th Joint Conference of Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Stroudsburg, PA: Association for Computational Linguistics, 2009: 1003-1011.

- [26] Surdeanu M, Tibshirani J, Nallapati R, *et al.* Multi-instance multi-label learning for relation extraction [C]// Proc of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Stroudsburg, PA: Association for Computational Linguistics, 2012: 455-465.
- [27] Hendrickx I, Kim S N, Kozareva Z, *et al.* Semeval-2010 task 8: multi-way classification of semantic relations between pairs of nominals [C]// Proc of Workshop on Semantic Evaluations: Recent Achievements and Future Directions. Stroudsburg, PA: Association for Computational Linguistics, 2009: 94-99.
- [28] Socher R, Huval B, Manning C D, *et al.* Semantic compositionality through recursive matrix-vector spaces [C]// Proc of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Stroudsburg, PA: Association for Computational Linguistics, 2012: 1201-1211.
- [29] Hashimoto K, Stenetorp P, Miwa M, *et al.* Task-oriented learning of word embeddings for semantic relation classification [C]// Proc of the 19th Conference on Computational Natural Language Learning. 2015: 268-278.
- [30] Zeng Daojian, Liu Kang, Lai Siwei, *et al.* Relation classification via convolutional deep neural network [C]// Proc of the 25th International Conference on Computational Linguistics. 2014: 2335-2344.
- [31] Nguyen T H, Grishman R. Relation extraction: perspective from convolutional neural networks [C]// Proc of the 1st Workshop on Vector Space Modeling for Natural Language Processing. 2015: 39-48.
- [32] Lin Yankai, Shen Shiqi, Liu Zhiyuan, *et al.* Neural relation extraction with selective attention over instances [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 2124-2133.
- [33] Zhou Peng, Shi Wei, Tian Jun, *et al.* Attention-based bidirectional long short-term memory networks for relation classification [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 207-212.
- [34] 万静, 李浩铭, 严欢春, 等. 基于循环卷积神经网络的关系抽取方法研究 [J/OL]. 计算机应用研究: 1-6 [2019-02-22]. <http://kns.cnki.net/kcms/detail/51.1196.TP.20181225.1615.003.html>. (Wan Jing, Li Haoming, Yan Huanchun, *et al.* Relation extraction based on recurrent convolutional neural network [J/OL]. Application Research of Computers: 1-6 (2018-12-26) [2019-02-19]. <http://kns.cnki.net/kcms/detail/51.1196.TP.20181225.1615.003.html>.)
- [35] Cai Rui, Zhang Xiaodong, Wang Houfeng. Bidirectional recurrent convolutional neural network for relation classification [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 756-765.
- [36] Lin Yanka, Liu Zhiyuan, Sun Maosong. Neural relation extraction with multi-lingual attention [C]// Proc of the 55th Annual Meeting of the Association for Computational Linguistics. 2017: 34-43.
- [37] Zheng Suncong, Hao Yuexing, Lu Dongyuan, *et al.* Joint entity and relation extraction based on a hybrid neural network [J]. Neurocomputing, 2017, 257 (9): 59-66.
- [38] Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 1105-1116.
- [39] Li Qi, Ji Heng. Incremental joint extraction of entity mentions and relations [C]// Proc of the 52nd Annual Meeting of the Association for Computational Linguistic. 2014: 402-412.
- [40] Zheng Suncong, Wang Feng, Bao Hongyun, *et al.* Joint extraction of entities and relations based on a novel tagging scheme [C]// Proc of the 55th Annual Meeting of the Association for Computational Linguistics. 2017: 1227-1236.