文章编号:1673-0062(2018)01-0066-07

中文关系抽取技术研究

阳小华,张硕望,欧阳纯萍

(南华大学 计算机学院,湖南 衡阳 421001)

摘　要: 在网络数据杂乱繁多的今天,信息抽取越来越受到重视,而关系抽取作为信息抽取的一个重要研究方向也倍受学者们的关注.在对关系抽取的发展历史进行回顾的基础上,阐述了近五年里关系抽取研究中的主要方法和思路.根据关系抽取中所采用的不同技术,对比分析了他们在模型思路及技术实现上的异同,同时对优势和劣势比较了分析.最后,对关系抽取任务中存在的难点问题进行了阐述,并提出了未来可能的解决思路,旨在为关系抽取技术的进一步发展和应用提供有效的说明和借鉴.

关键词: 关系抽取; 模式匹配; 机器学习

中图分类号: TP391.1　　**文献标志码:** A

DOI:10.19431/j.cnki.1673-0062.2018.01.012

A Comprehensive Review on Relation Extraction

YANG Xiao-hua,ZHANG Shuo-wang,OUYANG Chun-ping

(School of Computer,University of South China,Hengyang,Hunan 421001,China)

Abstract: Entity Relation Extraction is an important part of Information Extraction,and it is paid more and more attention in the chaotic network data.On the basis of reviewing the development history of relation extraction,the main methods and ideas of the technical literature in relation extraction research are analyzed and expounded in recent five years. According to the different technology adopted by relation extraction task,this paper discusses the model and implementation method of relation extraction,and compares the strengths and weaknesses of each method.Then it points out the problem of mainstream methods and analysis of possible solutions of future,in order to provide an effective illustration and reference for the further development and application of relational extraction.

key words: relation extraction; pattern matching; machine learning

收稿日期: 2018-01-08

基金项目: 国家自然科学基金资助项目(61402220; 61502221);湖南省哲学社会科学基金资助项目(14YBA335; 16YBA323);湖南省教育厅科研资助项目(16C1378);南华大学研究生科研创新资助项目(2016XCX15)

作者简介: 阳小华(1963-),男,教授,博士生导师,主要从事大数据检索与挖掘、舆情监测等方面的研究.E-mail: 875153468@qq.com

0 引　言

随着互联网技术的蓬勃发展,万维网的文本信息成倍增长,迅速成为一个巨大的信息资源库,如何从网络中的无结构文本信息中提取出有用知识越来越受到关注,信息抽取(information extraction,IE)技术应运而生.关系抽取(relation extraction,RE)的任务是从大量的无规则语料中识别并获取实体之间的语法或者是语义上的关系,是信息抽取的重要研究课题之一,被广泛的应用在机器翻译,问答系统,知识图谱等多个研究领域当中.

关系抽取通常是指在已知文本中实体对的情况下,抽取实体间的命名关系,并将抽取出来的实体对和关系进行规范化表示,其一般的形式化描述为三元组的形式<E1,Rel,E2>,E1 与 E2 代表实体,Rel 代表实体间的关系.通过提取实体间关系,获取更多实体间的语义联系,可以帮助计算机更好的处理大规模网络文本数据,以及理解非结构化文本的语义信息,在自然语言处理领域具有广阔的应用前景.围绕着实体间关系的发现和抽取,不少学者都展开了深入的研究.本文从基于模式匹配的关系抽取,基于机器学习的关系抽取和基于混合方法的关系三个方面对实体关系抽取的相关研究行了综述性分析.

1 关系抽取的发展历程

关系抽取研究是通过信息理解会议(message understanding conference,MUC)[1],内容自动抽取会议(automatic content extraction,ACE)以及文本分析会议(text analyis conference,TAC)[2]等评测会议的推动下逐渐展开的.20 世纪 80 年代末以来,美国高级研究计划局主持召开了 MUC,MUC 一共举行了 7 届,为信息抽取指定了任务和评测体系,推动了信息抽取研究的发展,而关系抽取任务在 1998 年的 MUC—7 会议上首次提出,任务是抽取实体对之间存在的位置、职位、和产品等三大类别的关系.

MUC 会议于 1998 年停办,美国国家标准技术研究院(national institute of standards andtechnology,NIST)组织的 ACE 于 1999 年继续开始进行信息抽取方面的评测,ACE 的目标是关注新闻领域的实体和实体关系抽取,并为抽取任务提供了评测语料和实体关系类型,标志着关系抽取研究开始进一步细化.从 2009 年开始 ACE 被归入了 TAC 中的一个专题,关系抽取成为了知识库构建(knowledge base population)任务的重要组成部分.近几年,随着互联网大数据分析的兴起,关系发现任务也开始偏重网络上开放数据间的实体关系抽取了.在近几年的一些顶级会议上开放关系抽取也成为了关注的研究热点.当然,即使是面向网络大数据的开放关系抽取,也需要领域知识的辅助,因此目前实体关系抽取实质上是领域关系抽取与开放数据关系抽取的组合任务.

实体关系抽取通常采用准确率 P(precision)、召回率 R(recall)和 F 值来衡量抽取方法的性能.其计算表达式如下:

$$P = \frac{\text{正确抽取的关系实例个数}}{\text{抽取出的关系实例个数}}$$

$$R = \frac{\text{正确抽取的关系实例个数}}{\text{测试集中包含的关系实例总数}}$$

$$F = \frac{2 \times P \times R}{P + R}$$

2 关系抽取的主要技术

得力于评测会议的发展,关系抽取研究进展迅速,并研究得出了大量不同的模型和方法.这些关系抽取方法大体上可以分为基于模式匹配的方法,基于机器学习的方法,和基于混合模型的方法.基于模式匹配的方法由专家预先依据限定领域实体关系的语法或语义信息人工制定出一套规则模板,并把测试文本中与模板语义相匹配的关系实例抽取出来.基于机器学习的方法利用各种统计学习的算法如支持向量机算法(support vector machine,SVM)和条件随机场(conditional random fields,CRF)等,将关系抽取看作一个分类问题,从关系训练样例中抽取出特征进行学习,自动从测试语料中抽取关系.

2.1 基于模式匹配的方法

最初的关系抽取方法大多基于模式匹配,由于需要语言学家对需要抽取的领域做深入的了解和分析,穷举所有关系模板,使得效率低下,一些学者开始尝试研究抽取模式的自动生成方法,郑家恒[3]计算模式实例之间的相似度,采用单链法聚类抽取模式,并将模式一一分类,归并同一类型的模式实例并获得最终的关系抽取模式,但是聚类方法在不同领域的抽取效率还没有具体的定量标准;姜吉发[4]提出了基于自举的关系与关系模式获取方法(bootstrapping relational pattern acquisition method,BRPAM),通过初始的少量二元关

系种子,从大型语料集中抽取更多二元关系,但是所有二元关系均由初始的种子匹配而来,存在着数据的稀疏性问题.

虽然基于模式匹配的方法存在一些不足,但是对机器学习方法的研究和发展起到了很好的借鉴作用,并且在很多领域之中,数据的规模和质量制约着系统性能,使得模式匹配的方法仍然起着不可替代的作用.

2.2 基于机器学习的方法

Miller[5] 等在 2000 年设计出一个基于规则的上下文无关的信息分析器,并利用宾州树库来训练句法参数,并在 MUC-7 上的模板关系抽取中取得了较好的效果,证明了机器学习对于关系抽取任务的可行性.机器学习的方法相对模式匹配在关系抽取任务中有着显著的优点:不需要人工构建模板,自动学习语料,可移植性强于基于模式匹配的方法等,很多专家学者开始选择使用机器学习的方法抽取实体关系,并针对机器学习里的一些难点采用了不同类型的机器学习方法进行解决.基于机器学习的关系抽取根据人工参与度分为有监督、半监督和无监督三种机器学习方法.

2.2.1 基于有监督的机器学习

有监督的机器学习将关系抽取看作二元分类问题,人们使用人工标注语料得到正例和反例,通过语义分析、句法分析等,选取特征集合,构造合适的分类器并训练得到分类模型,常用的如条件随机场(support vector machine,SVM)和最大熵分类器,然后根据分类模型构造抽取器从无结构文本中抽取出实体关系,通常的抽取模型有条件随机场模型(conditional random field,CRF).

有监督的机器学习方法可以分为基于特征向量的方法的和基于核函数的方法.

在基于特征向量的机器学习方法里,董静[6] 等人分析了特征选择的方法,并对包含和非包含关系分开进行提取,并针对性分别添加了子类框架特征,先祖成分以及实体到依赖动词路径等特征,并在报纸等新闻数据上进行实验,F 值为 65.76%.甘利新等[7] 认为董静提出的依赖动词特征存在 2 个问题:1) 选择距离较后的实体最近的动词并不完全可靠;2) 依赖动词特征并非都能探测出实体间关系有无或类型.因此作者提出了最近句法依赖动词特征,然后在旅游景点数据集的实体关系抽取任务中的 F 值高于董静的结果.张苇如[8] 利用知网的符号义原挖据潜在的实体对,同时与 Wikipedia 上的实体进行映射生成句子实例,保证了句子准确性,提高了语料质量.章剑锋[9] 将关系抽取应用到观点挖掘中,并列举了一个句子中所有的实体对的组合,解决了指代消解的问题,同时将评价词周围的程度副词加入特征当中,获得了 66.16%的 F 值,表明程度副词可以提高主观性关系抽取的效率.高俊平[10] 把关系抽取看作序列标注问题,利用深层句法分析构造领域知识演化抽取模式,并利用条件随机场模型学习模式特征和标注句子成分,抽取领域知识演化关系,作者希望通过机器学习自动分类符合演化关系模式的简单语句和不符合演化关系模式的复杂语句,但是简单的机器学习算法并不能准确的分辨简单模式和复杂模式,最终作者选择了人工判别,得到的简单模式的 F 值为 77.15%,复杂模式的 F 值为 88.26%,获得了较好的结果.

特征是有监督机器学习的研究重点之一,有监督的机器学习需要利用大量特征,不适合处理高维的情况,因此很多研究选择了使用核函数的方法来克服基于特征向量方法的局限性,进行实体关系的抽取.

基于核函数的方法生成句子结构树并计算树与树之间的相似度,然后利用支持核函数的分类器算法进行关系抽取.2003 年,Zelenko[11] 首次将核应用到关系抽取领域当中,他提出一种定义在浅层语法分析树上的核,训练 SVM 分类器,在新闻语料中取得了较好的结果.Culotta[12] 扩展了 Zelenko 的工作,提出了一种增广依赖关系树来增强树核函数,使用 SVM 进行分类,增加了抽取结果的准确率.黄瑞红[13] 通过实验研究了卷积树核以及改良的最短依赖路径核对关系抽取的有效性,并根据结果判定单纯的最短依赖路径核不能有效的提高关系抽取性能.陈鹏[14] 认为不同的核函数在不同领域里的关系抽取任务上有不同的效果,通过添加实体句法树信息,依存信息为特征,将各种核函数进行加权求和,确定最优的凸组合核函数.并在基于旅游领域文本数据集的实验中 F 值为 62.9%.郭剑毅[15] 使用基于多核融合的实体关系抽取方法,对多项式中符合平面核函数以及卷积树核等进行加权融合,最终得到了比单一核函数更好的结果.虞欢欢[16] 利用卷积树核,在关系实例中加入了小类、大类属性和 GPE 角色等特征,构造可以捕获结构化信息和实体语义信息的合一句法与实体语义关系树,有效的提高了关

系抽取的性能.李妩可[17]针对实体关系抽取方法中忽略了关系特征序列之间的模式差异,利用现有的语义词典和机器学习算法进行实体特征词标识并进行消歧,并根据提取模式的重要性设定权值加入到基于语义序列核函数的相似度计算中,最终获得了比单纯基于语义序列核函数更好的结果,该方法同时还具有不错的泛化能力.

基于核函数的方法可以挖掘出关系语料的深度特征,利用多种不同的数据组织形式表示实体关系,但是核函数计算复杂,时间花销较于基于特征向量的抽取方法更大,不适合处理大规模的关系抽取任务.

2.2.2　基于半监督的机器学习

在基于半监督的机器学习中,很多基于自学习的弱监督方法,利用少量关系实例种子进行机器学习,在迭代的过程中,不断将抽取出的高质量的关系实例加入到种子中,扩充新的训练语料,达到实体关系抽取效果.

针对词法特征矩阵的稀缺性问题,陈立玮[18]提出了 n-gram 特征用于缓解传统词法特征的稀疏性问题,并利用 bootstrapping 方法通过每次迭代都引入一部分高置信度样本扩充训练集中的正例,强化了抽取模型,最终得到了准确率的上升,但是召回率不升反而有所下降,其原因在于 n-gram 特征过于严厉使得大量可靠实体对被标为了负例,降低了关系抽取的召回率.贾真[19]在陈立伟的基础上,使用 n-pattern 模式代替 n-gram 模式,该模式放宽了 n-gram 的取词限制,解决了 n-gram 模式本身存在的问题,并通过最终的实验结果发现 2-pattern 的效果最好.黄卫春[20]将共现句分为简单关系句和复杂关系句,并利用模板匹配得到实体对,根据实体对和初始关系元组共现句的上下文位置的信息增益值相比对,得到符合要求的实体对,同时利用知网和同义词词林对关系描述词进行了扩展,有效提高了召回率.

半监督学习的方法可以有效减少人工参与的程度,但是,通过迭代自动获取的数据含有大量的噪声且存在语义偏移的情况,如何更加精准的分类和除噪,提高分类器的性能是当前基于半监督的机器学习方法的研究重点.

2.2.3　基于无监督的机器学习

基于无监督的关系抽取方法多采用模式聚类的方式,不需要事先对关系类型进行定义,无需人工对语料进行标注,可以避免由于关系模式建立不全而遗漏实体关系的问题,且可移植性较好.部分学者开始尝试使用无监督的机器学习方法提取实体关系.

无监督关系抽取方法由 Hasegawa[21]首次提出,通过对包含命名实体对的文本进行聚类,使用其结果来表示关系类型,在新闻领域语料上取得了一定的成果.马超[22]在领域本体构建任务中发现由于概念关系过于复杂使得人为标注费时费力,于是在 Hasegawa 的抽取方法基础上,引入了带样例概念关系对权重的无监督关系抽取算法,在聚类迭代过程中会计算样例实体对的权重并引用到下一轮迭代的综合特征上,其最终结果相比传统无监督方法较好,但是对抽取出的关系的置信度判别上还需要更多的研究以提高召回率.贾真[23]等在开放文本的部分-整体关系抽取问题上,利用 K 元模式提取算法提取实体对和实体对模式并进行协同聚类,使用 L1 正则化逻辑回归模型选择特征并进行模式匹配得抽取实体关系,得到了较好的召回率、F 值,但是准确率还有提升空间.

无监督的机器学习经常应用在开放领域的实体关系抽取当中,无需事先规定关系类型,但是前提是需要一个大规模的语料库用于挖掘关系模式,难点在于如何获取高置信度的模式,错误的模板会影响实体关系抽取的准确度.

2.3　基于混合的方法

无论是机器学习的方法还是模式提取的方法,都有他们的优势和不足,如何将不同的方法结合起来,利用其优势,取得更好的结果成为一些专家学者的研究目标.

于东[24]在抽取人物职衔履历属性中使用了模式匹配结合机器学习的方法,实现了字符串模式和依存模式的协同挖掘,由于候选集的噪声问题,最终的 F 值为 55.37,仍有很大改进空间.林如琦[25]将特征向量和卷积树核函数进行结合构造了一个混合模型,通过对特征相似度和树核相似度的加权融合得到一个综合相似度,对实体关系进行抽取,最终结果的准确度和 F 值均比单一方法的结果更高.黄晨[26]结合了卷积树核和模式聚类的方法,利用卷积树核计算结构化信息的相似度,然后进行实体聚类,并将相似实体归为相同簇来实现实体关系抽取.最终得到的 F 值比基于特征向量的无监督机器学习高出 3 个百分点,但是与有监督的方法比仍有一定的差距,需要进一步研究以获取更有效的平面特征来表征关系实例.

针对不同方法的局限性,合理的混合多种方

法可以有效的利用各方面的优势,目前需要更多的研究来发现更好的结合方法,提高效率和精度.表 1 是不同关系抽取方法的优劣势对比.

表 1 各关系抽取方法的优势与劣势

Table 1 Advantages and disadvantages of each relationship extraction method

方法	人工需求	领域移植性	所需时间	适合大小任务
模式匹配	高	弱	长	小
特征向量	较高	较弱	短	小
核函数	较高	较弱	较长	小
弱监督	较低	较强	较短	小
无监督	低	强	较短	大

3 目前关系抽取的主要难点

3.1 特征的质量不一

特征是基于机器学习的实体关系抽取算法的关键之一,一个分类器是否优秀依赖于当前特征对不同类型实例差异的表现能力.当前经过实践公认效果较好、使用相对频繁的特征如表 2.

表 2 特征空间

Table 2 Feature space

特征	说明
词法特征	包括实体类型,实体词、实体词性以及实体间词距等上下文信息及位置信息.
句法特征	依存句法关系,最短路径包含树,依赖动词等信息.
语义特征	包括实体属性信息,语义角色信息等.

当前特征存在的不足在于: 1) 常用特征已经饱和,如何自动挖掘新的高质量特征是急需解决的问题; 2) 特征在不同领域中的表现不一,无效或表达能力差的特征的增多导致特征空间的维度过高,造成时间开销的增大,效率下降.例如陈立玮[18]在抽取海量网络关系时,在实验中发现词法特征会因为该领域的句子的修饰成分过于具体,很难在其他句子中再次出现,因而无法做出贡献,针对这一弊端,作者选择了摒弃词法特征.

传统的词法特征以及语义特征已经趋近饱和,而基于深度分析的句法特征还具有一定的潜力,刘丹丹[27]以《同义词词林》为例,在基于核函数的关系抽取中加入了词汇语义特征验证抽取效果; 段利国[28]在刘丹丹的基础上,提出了同时利用《同义词词林》知网的语义编码树提取实体词信息作为特征.毛小丽[29]引入信息增益,期望交叉熵等文本分类里的特征选择算法来去除无用特征,李艳玲[30]利用最大熵和 SVM 双模型投票提高抽取效果.以上方法提示了知网以及同义词词林等语义词典以及未来大型知识库的词属性等信息或可成为更高效的特征来源,同时从信息增益的角度考虑了特征的去糟,对当前特征问题的解决起到了一定的提示和引导作用.

3.2 缺少有效训练语料

训练语料是基于机器学习的实体关系抽取算法的另一个关键之一,基于机器学习的分类器需要足够的训练语料进行学习才能达到好的分类效果,训练语料的不足以及关系实例在多个类别上分布不均匀都会导致准确率和召回率偏低.

当前训练语料存在的问题在于: 1) 人工标注训练语料费时费力; 2) 对关系类型的覆盖参差不齐.

在目前针对该类问题的解决方法中,Mintz[31]等人提出了远程监督(distantly supervised)方法,利用现有知识库对齐实体对抽取关系.杨宇飞[32]利用互动百科的 infobox 信息抽取学校领域的半结构化实体关系,经过贝叶斯算法优化后作为扩充训练语料.薛丽娟[33]通过人工设定规则推理引擎推理扩充训练语料.Savenkov[34]从社区问答知识库(community question answering,CQA)中抽取实体关系.

由此可看出目前针对训练语料匮乏的解决办法多为依赖外部网络知识库自动抽取,该方法十分依赖知识库的完备性和准确性,且在匹配的过程中容易产生噪声影响训练数据的质量,先抽取再除噪的模式的方法效果有限,如果能在抽取之前设定规则减少或消除噪声的产生则可以有效缓解训练语料不足带来的问题.

3.3 关系类型需要预先定义

关系类型是关系抽取任务的核心,传统的关系抽取发生在限定文本领域、限定语义单元类型的条件下,需要事先人为定义关系类型,而人工定义关系类型难免会出现人为定义不准确,类型定义不完全等问题,在前 web 数据信息海量增长的环境下显得越来越急需解决.

对于此,有专家学者提出开放式信息抽取(open information extraction,OIE).其成果主要有

Yates[35] 等设计实现的 TextRunner 系统、Wu[36] 等实现的 WOE 关系抽取系统以及 Fader[37] 等提出的 ReVerb 系统.此类系统均无需定义关系类型,自发抽取所有可能类型的关系,这种方式取得了一定的成就,但只能以抽取动词为核心的关系三元组为主,忽略了相对少量的名词形容词关系三元组,且其中利用依存分析通过核心动词匹配实体对十分依赖依存分析工具的准确性,目前技术成熟的中文依存分析工具主要有哈工大社会计算与信息检索研究中心研发的"语言技术平台(language technology platform,LTP)"以及斯坦福大学研究所的 Stanfor Parser 等,此类分析工具的劣势在于对于中文复杂长句的分析结果的准确度相对偏低,需对长难句进行一些预处理操作,如分句,主语补全等来缓解该问题.基于开放领域的关系抽取 F 值目前在 70%左右,与相对更成熟的领域内关系抽取的结果相比还有一定差距,需要进一步的探究.

4　结　论

本文阐述了关系抽取研究的发展历程,从技术方法角度分析了近五年最新的关系抽取相关文献,对比了各类方法的优点与不足.

关系抽取研究从最初基于规则设定,再到机器学习方法的引入,减少了人工参与的工作量,提高了抽取效率; 为了进一步的缩减人工参与工作量,专家开始研究基于半监督和无监督的机器学习,对标注语料的依赖性大大降低,但是训练语料质量问题导致其准确率和召回率不及基于模式匹配的方法和有监督机器学习的方法; 而随着关系类型的增多以及关系抽取范围的增加,关系抽取研究领域从限定领域开始向开放 web 领域扩展,开放式关系抽取借助 web 中海量数据完成实体关系抽取,具有广阔的应用前景; TAC 会议后,关系抽取并入知识库构建任务当中,远程监督方法开始受到关注,利用关系抽取充实 web 中的大规模数据库,而大规模数据库又可以为关系抽取任务提供训练语料和极性判断,远程监督方法为关系抽取提供广阔的思路.

目前关系抽取领域的研究还在不断进行中,近两年研究成果中,大部分研究仍然以特定关系抽取为主,特别在人物关系领域以及医学研究领域能得到很好的效果,而在半监督关系抽取中,标签传播和协同训练的思想被多次引入,为获取高质量的训练语料起到了不可忽视的作用.随着关系抽取技术的发展,这些方法和思想也将继续研究下去,并对机器翻译,大数据分析,问答系统的产生深远的影响.

参考文献:

[1] Surhone L M, Tennoe M T, Henssonow S F, et al. Message understanding conference [M]. Echeveria: Betascript Publishing, 2010.

[2] Giannakopoulos G, El-Haj M, Favre B, et al. TAC 2011 MultiLing Pilot Overview. In Text Analysis Conference (TAC) 2011 [C] // Text Analysis Conference. Gaithersburg, Maryland, USA: National Institute of Standards and Technology, 2001.

[3] 郑家恒,王义飞,李飞.信息抽取模式自动生成方法的研究 [J].中文信息学报,2004,18(1):48-54.

[4] 姜吉发,王树西.一种自举的二元关系和二元关系模式获取方法 [J].中文信息学报,2005,19(2):71-77.

[5] MILLER S, FOX H, RAMSHAW L, et al. A novel use of statistical parsing to extract information from text [C] // Association for Computational Linguistics, North American Chapter of the Association for Computational Linguistics Conference. Seattle Washington: Association for Computational Linguistics, 2000: 226-233.

[6] 董静,孙乐,冯元勇,等.中文实体关系抽取中的特征选择研究 [J].中文信息学报,2007,21(4):80-85.

[7] 甘利新,万常选,刘德喜,等.基于句法语义特征的中文实体关系抽取 [J].计算机研究与发展,2016,53(2):284-302.

[8] 张苇如,孙乐,韩先培.基于维基百科和模式聚类的实体关系抽取方法 [J].中文信息学报,2012,26(2):75-81.

[9] 章剑锋,张奇,吴立德,等.中文观点挖掘中的主观性关系抽取 [J].中文信息学报,2008,22(2):55-59.

[10] 高俊平,张晖,赵旭剑,等.面向维基百科的领域知识演化关系抽取 [J].计算机学报,2016,39(10):2088-2101.

[11] ZELENKO D, AONE C, RICHARDELLA A. Kernel methods for relation extraction [J]. The journal of machine learning research, 2003, 3: 1083-1106.

[12] CULOTTA A, SORENSEN J. Dependency tree kernels for relation extraction [C] //Meeting of the Association for Computational Linguistics, 21-26 July, 2004. Barcelona Spain: Association for Computational Linguistics, 2004: 423-429.

[13] 黄瑞红,孙乐,冯元勇,等.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.

[14] 陈鹏,郭剑毅,余正涛,等.基于凸组合核函数的中文领域实体关系抽取 [J].中文信息学报,2013,27(5):

145-148.

[15] 郭剑毅,陈鹏,余正涛,等.基于多核融合的中文领域实体关系抽取[J].中文信息学报,2016,30(1):24-29.

[16] 虞欢欢,钱龙华,周国栋,等.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):17-22.

[17] 李妩可,郭赛球,尹艳.命名实体关系抽取算法的改进[J].计算机工程,2010,36(24):289-290.

[18] 陈立玮,冯岩松,赵东岩.基于弱监督学习的海量网络数据关系抽取[J].计算机研究与发展,2013,50(9):1825-1835.

[19] 贾真,杨燕,何大可.基于弱监督学习的中文百科数据属性抽取[J].电子科技大学学报,2014,43(5):758-762.

[20] 黄卫春,徐力,熊李艳,等.基于信息增益的 Web 人物关系抽取[J].计算机应用研究,2016,33(8):2286-2289.

[21] HASEGAWA T,SEKINE S,GRISHMAN R.Discovering relations among named entities from large corpora [C]//Meeting on Association for Computational Linguistics.Association for Computational Linguistics.Barcelona Spain:Association for Computational Linguistics,2004:415.

[22] 马超.基于 Web 信息使用改进的无监督关系抽取方法构建交通本体[J].计算机系统应用,2015,24(12):273-276.

[23] 贾真,何大可,尹红风,等.基于无监督学习的部分-整体关系抽取[J].西南交通大学学报,2014,49(4):591-595.

[24] 于东,刘春花,田悦.基于远距离监督和模式匹配的职衔履历属性抽取[J].计算机应用,2016,36(2):455-459.

[25] 林如琦,陈锦秀,杨肖方.多信息融合中文关系抽取技术研究[J].厦门大学学报(自然科学版),2011,50(3):540-544.

[26] 黄晨,钱龙华,周国栋,等.基于卷积树核的无指导中文实体关系抽取研究[J].中文信息学报,2010,24(4):11-16.

[27] 刘丹丹,彭成,钱龙华,等.《同义词词林》在中文实体关系抽取中的作用[J].中文信息学报,2014,28(2):91-98.

[28] 段利国,徐庆,李爱萍,等.实体词语义信息对中文实体关系抽取的作用研究[J].计算机应用研究,2017,34(1):141-145.

[29] 毛小丽,河中市,邢欣来.基于特征选择的实体关系抽取[J].计算机应用研究,2012,29(2):530-532.

[30] 李艳玲,林民.基于双模型投票的人物关系抽取研究[J].计算机应用研究,2017,34(3):773-776..

[31] MINTZ M,BILLS S,SNOW R,et al.Distant dupervision for relation extraction without labeled data [C]//Mintz,Joint Conference of the Meeting of the ACL and the International Joint Conference on Natural Language Processing of the Afnlp:Volume.Suntec Singapore:Association for Computational Linguistics,2009:1003-1011.

[32] 杨宇飞,戴齐,贾真.基于弱监督的属性关系抽取方法[J].计算机应用,2014,34(1):64-68.

[33] 薛丽娟,席梦隆,王梦婕.基于规则推理引擎的实体关系抽取研究[J].计算机科学与探索,2016,10(9):1310-1320.

[34] SAVENKOV D,LU W L,DALTON J,et al.Relation extraction from community generated question - answer pairs [C]//Conference of the North American Chapter of the Association for Computational Linguistics:Student Research Workshop.New Orleans Louisiana:Association for Computational Linguistics.2015:96-102.

[35] YATES A,CAFARELLA M,BANKO M,et al.TextRunner:open information extraction on the web [C]//Human Language Technologies:the Conference of the North American Chapter of the Association for Computational Linguistics:Demonstrations.Rochester New York:Association for Computational Linguistics,2007:25-26.

[36] WU F,WELD D.autonomously semantifying wikipedia [C]//Sixteenth ACM Conference on Conference on Information and Knowledge Management.ACM.New York:Association for Computational Linguistics,2007:41-50.

[37] FADER A,SODERLAND S,ETZIONI O.Identifying relation for open information extraction [C]//Conference on Empirical Methods in Natural Language Processing,EMNLP 2011.Edinburgh,United Kingdom,2011:1535-1545.

(责任编辑:扶文静)