面向特定领域文本的重叠关系语料库构建方法

刘 凯,廖湘琳,张宏军
(陆军工程大学 指挥控制工程学院,江苏 南京 210000)

摘 要:实体关系语料库是信息抽取领域的基础数据资源,其规模和质量直接影响信息抽取深度学习模型的效果。目前建立的特定领域语料库在重叠关系方面的研究较少,且现有方法需要高昂的人工标注成本。该文融合已有的基于实体识别和触发词规则的语料标注算法,基于自定义关系 schema 实现网络文本中重叠关系的自动标注。首先,借助特定领域专业词典进行命名实体识别,构造命名实体集;然后根据自定义关系模式 schema 和依存句法分析进行特征词聚类,构造触发词词典;最后,基于命名实体集和触发词词典进行语料回标。该算法有效减少了人工标注量,标注速度快,标注后的语料规模较大,有效提取重叠关系信息,为特定领域信息抽取扩充语料库提供了可行方案。同时,该文探讨了数据源可用性,评价了标注质量并对语料库进行了统计分析。实验结果显示,该方法总体回标成功率为 76.7%,总体关系标注准确率为 85.8%,利用基础重叠关系抽取模型进行实验,实验结果 F1 值达到 93.68%。

关键词:实体关系;信息抽取;语料库构建;schema;触发词
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2022)10-0126-06
doi:10.3969/j.issn.1673-629X.2022.10.021

Constructing of Corpus of Overlapping Relationships for Domain-specific Text

LIU Kai,LIAO Xiang-lin,ZHANG Hong-jun
(School of Command and Control Engineering,Army Engineering University of PLA,Nanjing 210000,China)

Abstract:The corpus of entity relations is the basic data resource in the field of information extraction,and its scale and quality directly affect the training effect of the deep learning model. There is little research on overlapping relationships for domain-specific corpus at present,and existing methods require high manual annotation cost. We incorporate the existing annotation algorithm based on entity recognition and trigger word rules and implement the automatic annotation of overlapping relations in network text according to the custom relations schema. First,after the named entities were identified by professional dictionary in the specific field,the named entity set was constructed. Then the trigger word dictionary was constructed by clustering the feature words according to the custom relational pattern schema and dependency parsing. Finally,the corpus automatic annotation was carried out based on the named entity set and the trigger word dictionary. The proposed algorithm can effectively reduce the amount of manual annotation,with fast annotation speed and large scale of corpus after annotation,which extracts the information of overlapping relations effectively and provides a feasible scheme for expanding the corpus in information extraction of specific field. Meanwhile,we explore the availability of data source,evaluate the quality of annotation and make statistical analysis of the corpus. The experimental results show that the overall success rate of the proposed method is 76.7%,the overall relationship annotation accuracy is 85.8%. In the experiment using the basic overlap relations extraction model,the value of F1 reaches 93.68%.

Key words:entity relations;information extraction;corpus construction;schema;trigger word

0 引 言

现如今,信息抽取领域中,以给定关系模式(schema),通过有监督学习方式对深度学习模型进行训练,进而完成对文本数据信息抽取的过程已被广泛应用,且在重叠关系三元组抽取上有明显效果[1-2]。该文对爬取得到的新闻网络文本进行分析,发现特定领域实体间存在重要的重叠关系,但是受标注语料匮乏问题的制约,信息抽取领域在这方面的研究较少,无

收稿日期:2021-11-12 **修回日期**:2022-03-16
基金项目:国家自然科学基金(61806221)
作者简介:刘 凯(1996-),男,硕士研究生,研究方向为自然语言处理;通信作者:张宏军(1963-),男,教授,博导,博士,研究方向为军事仿真、数据工程、效能评估。

法满足国内外研究者们的需求。所以,为了更高效准确地抽取文本中的重叠关系,该文构建重叠关系标注语料库,为信息抽取模型训练提供丰富数据,为当前国内信息抽取语料库构建和完善提供借鉴和参考。

如何完善地构建实体关系模式,如何高效准确地构建特定领域实体重叠关系抽取标注语料库是该文的研究重点。目前网页新闻和网络博客等开放领域是大部分语料库的主要数据来源,如公开的中英文关系抽取语料库 DuIE1.0[3],其关系类型主要包含常见的人物关系,CMeIE[4] 为医学领域关系语料库, FewRel[5] 关系数据集包含多领域的关系类型。该文借助远程监督知识,依据命名实体识别、依存句法分析和触发词词典,基于自定义 schema 对网络文本中的重叠关系进行语料标注,构建关系抽取语料库。其主要描述作战力量编成部署信息。

1　相关工作

语料库构建工作,过程复杂,形式多样。针对通用语料库的构建工作已经取得很多成果。比如周惠巍等人[6]依据词性和句子结构等信息构建中文模糊限制信息语料库,为事件信息抽取提供资源支持。蒋贻顺[7]构建触发词词典,通过规则匹配实现人物关系三元组抽取。针对特定领域的研究,目前更多的研究集中在地理实体关系[8-9]、医学领域[10-11]和军事领域。苟继承[12]利用远程监督方法,基于规则匹配的方式获得实体关系信息,构建实体关系知识库。蒋序平等人[13]通过定义事件模板,构造触发词词典,形成人工标注种子数据集,经过模型训练迭代生成针对军事想定文本事件抽取的语料库。冯鸾鸾等人[14]制定了一系列标注规范,对收集到的海量互联网文本进行术语语料标注,并且构建出面向国防科技领域的技术和术语语料库。上述方法需要人工构建规则,增加人工标注负担。该文巧妙融合上述研究方法,根据自定义实体关系 schema 对收集到的特定领域新闻网络文本进行自动回标。该方法避免了大量的人工标注工作,构建出的语料库规模较大,质量较高,有较大实用价值。

2　语料库构建方法

面向特定领域文本的重叠关系抽取语料库构建流程如图 1 所示。

2.1　语料来源

该文将语料限定在特定领域内,为了发现网络文本中重叠关系信息,建立一个通用的、实体覆盖面更广的关系类型模式。通过网络爬虫抓取来自新浪网、光明网、国防科技信息网、武器百科大全网站等超过 1 000个网页,获得原始数据约 10 万条,占用空间资源 26.3 M。数据样例如下所示:(1)近日,北京武警放出了使用 QMK171 瞄准镜的 95-1 式的照片,意味着 QMK171 瞄准镜已经大量入役。(2)日前,美国通用动力公司在美国首都华盛顿举行的美国陆军协会年会博览会上展示了其最新的 RM277 型全自动轻机枪的信息,将采用美军最新研发的 6.8 毫米弹药,等等。

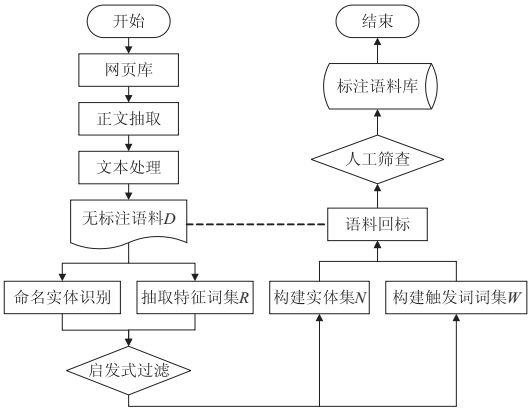

图 1　语料库构建流程

新闻类语料来源于网页。通过观察网页源代码中的 HTML 标签和文字分布特点,利用 python 的爬虫库 BeautifulSoup 解析网页源代码,对网页中正文较集中的内容块进行文本提取。

正文提取完成后,为方便后续实体关系的抽取,将语料数据进行分句处理。中文语句的一句话通常由句号“。”、问号“?”、感叹号“!”、省略号“……”等符号结尾,利用这些符号作为句子分割条件,得到分句后的无标注文本数据集 D,作为语料库构建的数据来源。

2.2　关系模式构建

ACE 评测会议于 2005 年公布了官方标注的关系抽取语料库,包括中文、英文、阿拉伯文的标注语料,其定义了表中的 6 类大类关系和 18 类小类关系的关系类型。COAE 会议于 2016 年针对中文领域关系抽取推出包含 10 种关系类型的中文关系抽取训练集。

但是上面两个数据集的关系体系与特定领域的关系具有一定差异,无法成为构筑特定领域关系体系的基础。通过专家知识和对特定领域文本的分析,根据上述关系分类,对实体关系的筛选,过滤与领域无关的大量内容,经过整理,该文最终预定义了 5 种命名实体,分别是组织(ORG)、武器(WEAP)、地点(LOC)、行动(ACT)、人员(PER);7 种实体关系类别,分别是人员和组织的隶属关系、组织与组织的编成关系、组织与行动的执行关系、组织与地点的部署关系、行动与地点的目标关系、组织与武器的配置关系。关系 schema 如下:

{"object_type":"ORG","predicate":"编成","subject_type":"ORG"}

{"object_type":"ACT","predicate":"执行","subject_

type": "ORG"}

{"object_type": "LOC", " predicate ": "部署", "subject_type": "ORG"}

{"object_type": "LOC", " predicate ":"布置", "subject_type": "WEAP"}

{"object_type": "LOC", " predicate ": "目标", "subject_type": "ACT"}

{"object_type": "WEAP", " predicate ":"配置", "subject_type": "ORG"}

{"object_type": "ORG", " predicate ": "隶属", "subject_type": "PER"}

通过分析语料文本,存在如图 2 中三种重叠关系,以此为基准进行下一步研究。

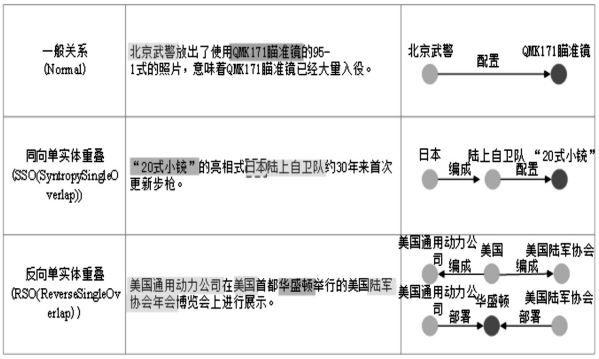

图 2 重叠关系示例图

2.3 基于自定义关系 schema 的重叠关系语料标注

2.3.1 实体集构建

根据 2.2 节中确定的五种实体进行以下分析:首先利用命名实体识别方法和自制的领域专业词典,将 2.1 节构建的训练语料输入 BiLSTM+CRF 命名实体识别模型[15]进行实体识别,然后通过启发式规则,比如去掉单字符名词、保留专有名词等进行人工筛选,最后获得备选实体集 N,为后续启发式实体关系对齐和关系数据去噪做准备。备选实体集 N 部分实体如表 1 所示。

命名实体识别所用标注数据集由多人进行手动标注并打分评估进行融合所得。

2.3.2 触发词词典构建

触发词词典构建过程为:首先进行特征词抽取(运用 LTP 工具抽取动词、名词),然后根据 schema 聚类成触发词词典,最后根据实体对进行启发式关系过滤。

(1)特征词抽取。

通过观察语料库发现,绝大多数产生关系的实体对都可以由其上下文中一般动词或者一般名词触发和描述(统称为特征词),而且这些特征词均与待处理的实体对在依存句法分析树中产生有限的几类关系。

表 1 备选实体集 N 部分实体示例

类别	部分实体示例
组织(ORG)	“美军”,“日本陆上自卫队”,“特种部队”,“中国陆军”,“国防部”,“联合参谋本部”,“海军军事学术研究所”,“法国海军”,“美军舰队”……
人员(PER)	“拉尔德”,“埃尔莫·祖沃”,“伊丽莎白”,“马沙尔中将”,“西田正雄大佐”,“田中赖三大佐”,“陈水扁”,“政委丁海春”,“海军副司令员张永义”……
地名(LOC)	“淮河”,“金刚山”,“加勒比海”,“关塔那摩湾”,“古巴”,“加勒比海”,“齐柏林”,“新泽西州”,“洛杉矶市”,“格鲁吉亚”,“伊拉克”,“莫桑比克”……
武器(WEAP)	“M249 轻机枪”,“RM277 型全自动轻机枪”,“步枪”,“6.8 毫米弹药”,“AR 步枪”,“预警直升机”,“轻型航空母舰”,“926 型潜艇支援舰”……
行动(ACT)	“作战”,“打击”,“巡逻”,“反恐”,“制裁”,“支援”,“防御”,“护航”,“空中掩护”,“救援”,“防御”,“维和”,“袭击”,“军事行动”,“军事冲突”……

特征词抽取过程是为了抽取语料库中与特定实体对类型下的实例共现,且依存句法分析后具有特定语义关系的动词或名词。然后采用启发式过滤规则,进行特征词集过滤筛选[16]。

词性分析和依存句法分析中,使用哈工大语言技术平台(Language Technology Platform,LTP)的处理模块。LTP 处理中文文本具有良好的性能。首先对语料库进行词性标注,抽取出动词或动名词。LTP 定义了 15 个依存句法标签,包括主谓关系(SBV)、动宾关系(VOB)、间宾关系(IOB)、并列关系(COO)等。

具体步骤如下所示:

①根据 2.2 节中构建的 schema,得到特定实体对类型的槽(socket)。对每个实体 $n \in N$,在语料 D 中检索包含实体的所有句子,保留那些同时包含实体 n_i 和另一个与其形成特定实体对类型的实体 n_j 的句子 Sent,由此形成七种关系句子集 $< n_i, n_j, \text{Sent} >_m (m = 1,2,\cdots,7)$。

②对 $< n_i, n_j, \text{Sent} >_m$ 中包含的所有句子进行词性标注,抽取所有动词和名词,按照如下启发式规则进行统计过滤,得到候选特征词集 R_m。

Rule1:根据依存句法分析后,动词或名词必须满足与实体对中任一实体存在主谓宾结构 SBV-VOB、从

属关系结构 ATT-ATT、动补介宾关系结构 CMP-POB。

③对于每一个 $w \in R$,统计其在第(1)步得到的句子 Sent 集中出现的频率 $P_S(w_k)$,去掉频率小于常数 θ 的特征词。

④根据候选特征词 w_k 在 D 中和特定实体对类型句子集 Sent 中的分布信息,采用以下公式计算其与实体对类型的相关度 $\mathrm{Rel}(w_k)$[16],其中 $P_S(w_k)$ 和 $P_D(w_k)$ 分别表示 w_k 在特定实体对类型句子集和语料库 D 中的频率。

$$\mathrm{Rel}(w_k) = P_S(w_k)/P_D(w_k)$$

⑤根据相关度对候选特征词进行排序,根据排序位置取靠前的 Top-K 个作为特征词,获得筛选后候选特征词集 R。

(2)Schema 聚类与触发词词典构建。

一系列具有相同含义和用法的特征词可以体现同一种关系,因此根据 2.2 节 Schema 中确定的七种关系词对上述包含七种关系类型的候选特征词集 R 进行对应聚类,构建触发词词典 W,如表 2 所示。

表 2　触发词词典部分触发词示例

关系类型 r	ID	部分触发词 w
编成关系	1	{下辖,麾下,近卫,进驻,纳入,隶属,改编,编有,建制,调拨,编成}
部署关系	2	{突破,海上,混编,遍及,协同,进攻,作战,补给,远洋,纵深,支援,跨,部署,巡逻}
配置关系	3	{装备,装设,布置,布局,配备,配置}
部置关系	4	{设置,装设,指挥台,置于,布设}
目标关系	5	{跟踪,追踪,精确,袭,捕捉,侦测,瞄准,隐藏,来袭,拦截,打击,识别,警戒,射击,引导,行进,实施}
执行关系	6	{训练,担负,遂行,侦察,作战,担负起,充当,胜任,遭敌,协同,担任,实施,承担,使命}
隶属关系	7	{调离,改编,前来,隶属于,麾下,受命,担当,调往,派往}

2.3.3　语料回标

借助实体识别和触发词规则,基于自定义关系 schema 的语料标注方法假设:如果训练语料的某一句话包含的实体集中的实体对在触发词词典中有对应的触发词,就认为这句话描述了触发词所表示的 schema 中的关系类型。基于此假设进行语料自动回标,有助于减少人工标注的工作量。

标注算法流程:首先,根据命名实体识别结果,获得实体和实体类型列表,然后顺序扫描根据领域词典进行结巴分词后的语料文本,依次匹配实体集中的实体,先进行头实体 subject 匹配,查找到一个实体后转为该文本片段尾实体 Object 匹配,然后根据 schema 槽中的实体对类型进行判断,两者是否相关,若相关,则提取关系信息,查找触发词词典,对关系类型标注和保存,否则继续进行实体匹配,此过程在句子集内循环,直到遍历完成单个句子中所有关系。此方法简单有效,标注效率高。算法如下所示:

算法 1:重叠关系语料回标算法。

输入:实体集 N,触发词词典 W,待匹配语料 D,schema;

输出:标注文本 s。

① for D 中的每一句话 s do:

② for 实体集 N 中的每一个实体和类型 type do:

③ 头实体[subject,s_type]匹配

④ if subject=匹配成功 then

⑤ for 实体集 N 中除 subject 外的每一个实体和类型 type do:

⑥ 尾实体[object,o_type]匹配

⑦ if object=匹配成功 then

⑧ if schema[s_type,o_type] and 对应关系 $r \leftarrow W[w]$ then

⑨ 标注文本←文本串 s +关系 r +subject+object

3　实验分析

为了保证语料库的专业性和可靠性,首先探讨本语料库数据源的可用性,然后进行标注质量评价并使用基础模型验证语料库的质量。

3.1　数据源可用性分析

对约 10 万条原始数据进行随机抽取,以评价新闻网站作为构建特定领域重叠关系抽取语料库的可用性。(1)从原始数据中随机抽取 1 000 条语句;(2)根据语句中包含的实体类型将其划分到文中的实体分类体系中;(3)统计每个实体类型下语句的信息量,结果如表 3 所示。

表 3　数据源可用性统计

分类	句子数	含三元组的句子数	关系三元组数	关系种类数
组织(ORG)	551	541	2 923	5
人员(PER)	233	132	1 104	1
地名(LOC)	305	196	1 514	3
武器(WEAP)	895	422	2 378	2
行动(ACT)	89	42	398	2
总计	2 073	1 333	8 317	7

由表 3 可以看到:(1)从原始数据中随机抽取的 1 000条新闻语句中最多有 89.5% 被成功划分到该文

提出的实体分类体系中,但是不同句子中,实体数量分布不均匀;(2)包含关系三元组的语句数约占抽取的句子总数的 64.3%,平均每个句子中含有 6 个关系三元组,涵盖了自定义的 7 种实体关系。可见通过新闻等网站爬取的原始语料蕴含了丰富的实体关系三元组,为构建实体关系语料库提供了充足的数据资源。

3.2 标注质量评价

基于数据可用性分析结果,从实体集 N 中分别为实体分类的 5 个实体类型选取 50 个实体,共计 250 个;然后对基于该方法构建的重叠关系语料库和实体识别筛选语料进行统计分析。特定领域的重叠关系语料库成功标注 18 750 个句子,占实体识别筛选语料的 51.3%。此语料库中的知识形式为{“text”:“文本”,“spo _ list”:“{subject, predicate, object}”},其中 subject 表示主语(头实体),object 是宾语(尾实体),predicate 是谓词(关系的抽象表示)。为了方便查询,依然采用 json 格式保存三元组信息,标注示例如:{"text":"海军军事学术研究所研究员里奇博士说:“这次建造轻型航母的决定是‘一石二鸟’,这将成为体现‘有效性的韩国海军核心战斗力’”。","spo_list":[{"predicate":"编成","object_type":"ORG","subject_type":"ORG","object":"海军军事学术研究所","subject":"韩国海军"},{"predicate":"配置","object_type":"WEAP","subject_type":"ORG","object":"轻型航母","subject":"韩国海军"},{"predicate":"隶属","object_type":"ORG","subject_type":"PER","object":"海军军事学术研究所","subject":"里奇"},{"predicate":"隶属","object_type":"ORG","subject_type":"PER","object":"韩国海军","subject":"里奇"}]}。

表 4 为数据统计信息。其中成功率表示成功匹配包含该实体的三元组的句子数占包含该实体的标注句子总数的百分比;准确率表示正确标注的三元组数占包含该实体的三元组数的百分比。实验中根据 250 个实体得到了实体识别筛选标注语料中的 1 024 条语句,通过随机抽样计算,语料库的总体回标成功率为 76.7%,总体关系标注准确率为 85.8%。

表 4 标注质量评价统计

类别	正确三元组数	标注三元组数	成功匹配句子数	标注句子数	成功率/%	准确率/%
组织(ORG)	2 425	2 850	539	569	94.7	85.1
人员(PER)	86	95	38	67	56.7	90.5
地名(LOC)	341	395	73	112	65.1	86.3
武器(WEAP)	187	204	47	89	52.8	91.7
行动(ACT)	193	223	88	187	47.0	86.5
总计	3 232	3 767	785	1 024	76.7	85.8

针对标注出的实体关系,进行如下统计展示。图 3 表示每句话中包含不同三元组数目的句子数;图 4 反映句子集中包含各类重叠关系的数目及三元组总数。

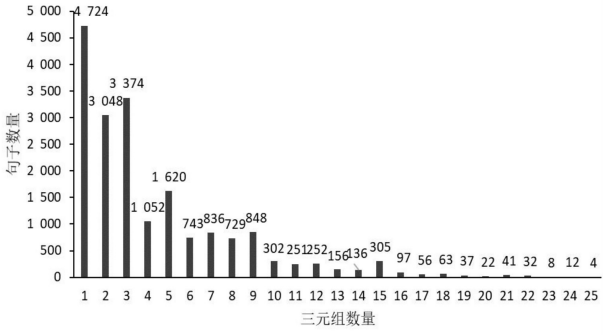

图 3 三元组频数统计

3.3 信息抽取模型实验

为了说明构建的语料库的可用性,实现对军事新闻中蕴含的作战力量编成部署信息的抽取,该文使用信息抽取基础模型 DGCNN + self – attention[17] 进行实验。将构建好的语料库按照 7 ∶ 3 的比例进行训练集和验证集的划分,并选择 17 942 条经过清洗后的语句作为测试集。评测采用传统的召回率(R)、准确率(P)、F1 值。模型实验结果显示,利用构建的语料库训练的基础模型,其准确率达到 95.98%,召回率达到 91.50%,F1 值为 93.68%,效果较好。

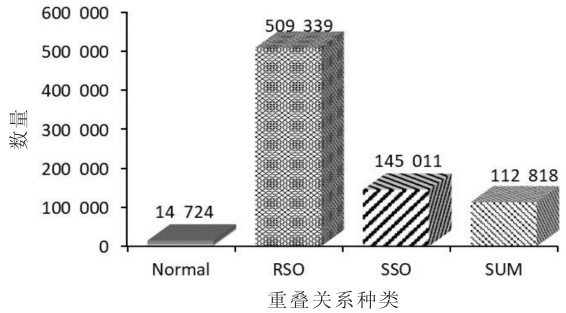

图 4 重叠关系频数统计

3.4 语料库结果可视化

为更好展示构建的语料库效果,采用 neo4j 图数据库存储并进行可视化。语料库部分语句各关系可视化如图 5 所示。

图 5　关系三元组可视化

以美国为例:如“美国”存在“美国-编成-美国特种作战司令部”、“美国-配置-黄蜂级航空母舰”、“美国-执行-护航”等三种关系,26 个关系三元组。

4　结束语

该文描述了面向特定领域文本的重叠关系抽取语料库构建工作。首先对爬取到的特定领域网络文本进行分析,构建关系模式 schema,然后利用命名实体识别模型对文本进行实体识别得到备选实体集,通过依存句法分析和特征词聚类构造触发词词典,最后基于实体集和触发词词典进行语料自动回标,构建出目前规模较大的面向特定领域的实体重叠关系抽取语料库。同时,探究了数据源的可用性和标注质量,语料总体的回标成功率为 76.7%,总体关系标注准确率为 85.8%,利用基础重叠关系抽取模型进行实验,实验结果 F1 值达到 93.68%。

文中的构建方法减少了人工标注的工作量,标注效率较快,质量较高。但是,由于网络文本的冗杂,构建的语料库仍存在部分实体和不常见实体无法识别,目标等关系数量相对较少,且包含的关系类型较少等问题。未来的工作中,将利用抽取模型进行迭代更新,改进标注质量,并且继续完善标注体系,扩大标注规模,为后续特定领域的信息抽取、知识图谱构建等工作奠定基础。

参考文献:

[1] WEI Z,SU J,WANG Y,et al. A novel cascade binary tagging framework for relational triple extraction[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. [s. l.]:Association for Computational Linguistics,2020:1476-1488.

[2] WANG Y,YU B,ZHANG Y,et al. TPLinker:single-stage joint extraction of entities and relations through token pair linking[J]. arXiv:2010.13415v1,2020.

[3] LI S,HE W,SHI Y,et al. DuIE:a large-scale Chinese dataset for information extraction[C]//Proceedings of CCF international conference on natural language processing and Chinese computing. Dunhuang:Springer,2019:791-800.

[4] GUAN T,ZAN H,ZHOU X,et al. CMeIE:construction and evaluation of Chinese medical information extraction dataset[C]//Proceedings of the CCF international conference on natural language processing and Chinese computing. Zhengzhou:Springer,2020:270-282.

[5] GAO T,HAN X,ZHU H,et al. FewRel 2.0:towards more challenging few-shot relation classification[C]//Proceedings of the international joint conference on natural language processing. Hong Kong,China:Association for Computational Linguistics,2019:6249-6254.

[6] 周惠巍,杨　欢,徐俊利,等. 中文模糊限制信息范围语料库的研究与构建[J]. 中文信息学报,2017,31(3):77-85.

[7] 蒋贻顺. 基于规则匹配与神经网络学习的中文实体关系抽取研究[D]. 合肥:合肥工业大学,2019.

[8] 王姬卜,陆　锋,吴　升,等. 基于自动回标的地理实体关系语料库构建方法[J]. 地球信息科学学报,2018,20(7):871-879.

[9] 陈振东. 基于领域适应迁移学习的地理实体关系抽取[D]. 武汉:武汉理工大学,2019.

[10] 刘一斌. 中医中文电子病历命名实体语料库构建及研究[D]. 广州:广州中医药大学,2020.

[11] 咎红英,刘　涛,牛常勇,等. 面向儿科疾病的命名实体及实体关系标注语料库构建及应用[J]. 中文信息学报,2020,34(5):19-26.

[12] 苟继承. 基于远程监督的军事实体关系抽取应用研究[D]. 成都:电子科技大学,2020.

[13] 蒋序平,战立莹,杨若鹏,等. 一种军事想定文本事件抽取语料库迭代式构建方法及装置:CN110597997A[P]. 2019.

[14] 冯鸾鸾,李军辉,李培峰,等. 面向国防科技领域的技术和术语语料库构建方法[J]. 中文信息学报,2020,34(8):41-50.

[15] LAMPLE G,BALLESTEROS M,SUBRAMANIAN S,et al. Neural architectures for named entity recognition[C]//Proceedings of the 15th annual conference of the North American Chapter of the association for computational linguistics. Stroudsburg:Association for Computational Linguistics,2016:260-270.

[16] 王莉峰. 领域自适应的中文实体关系抽取研究[D]. 哈尔滨:哈尔滨工业大学,2011.

[17] SU J. A hierarchical relation extraction model with pointer-tagging hybrid structure[EB/OL]. 2019. https://kexue.fm/archives/6671.