

面向低资源语言的实体关系联合抽取方法、装置、电子设备及存储介质

王磊^{*}; 马博^{*}; 杨振宇; 杨雅婷^{*}; 王震; 马荣; 周喜^{*}

2023-04-25

专利权人 中国科学院新疆理化技术研究所

专利类型 发明专利

摘要 本发明公开了一种面向低资源语言的实体关系联合抽取方法、装置、设备以及存储介质。该方法获取包含丰富资源语言和低资源语言的双语平行数据集，对句子对中的实体进行预处理；将处理得到的双语平行句子输入到XLM-Roberta多语言预训练模型中得到句子的特征表示；将实体的特征信息中通过注意力计算的方式添加相对位置信息；将双语实体对信息和双语句子对信息进行特征融合；通过二元分类器对融合得到的特征信息进行判断并通过判断结果构建伪标注数据；利用伪标注数据训练低资源语言的实体关系抽取模型。本发明通过相似度计算的方式跨语言的构建低资源语言伪标注语料并训练实体关系抽取模型。有效的弥补了低资源语言语料缺乏的问题。

申请日期 2022-12-26

申请号 CN202211673441.9

公开(公告)号 116011454A

代理机构 乌鲁木齐中科新兴专利事务所(普通合伙) 65106

文献类型 **专利**

条目标识符 <http://ir.tianshanzw.cn/handle/365002/9303>

专题 多语种信息技术研究室

推荐引用方式 王磊,马博,杨振宇,等. 面向低资源语言的实体关系联合抽取方法、装置、电子设备及存储介质. 116011454A[P]. 2023-04-25. **GB/T 7714**

条目包含的文件

条目无相关文件。

所有评论 (0)

[发表评论/异议/意见]

暂无评论

除非特别说明，本系统中所有内容都受版权保护，并保留所有权利。

个性服务

推荐该条目

- ★ 保存到收藏夹
- 查看访问统计
- 导出为Endnote文件

谷歌学术

- 谷歌学术中相似的文章
- [王磊]的文章
- [马博]的文章
- [杨振宇]的文章

百度学术

- 百度学术中相似的文章
- [王磊]的文章
- [马博]的文章
- [杨振宇]的文章

必应学术

- 必应学术中相似的文章
- [王磊]的文章
- [马博]的文章
- [杨振宇]的文章

相关权益政策

暂无数据

收藏/分享

