

**XJIPC OpenIR** > 多语种信息技术研究室

---

**融合词对齐适配器模块的跨语言文本表示方法、装置、设备及存储介质**

杨雅婷<sup>\*</sup>;艾合塔木江·艾合麦提;王磊<sup>\*</sup>;马博<sup>\*</sup>;董瑞;周喜<sup>\*</sup>

2023-03-10

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** 本发明公开了一种融合词对齐适配器模块的跨语言文本表示方法、装置、设备及存储介质,涉及人工智能、自然语言处理、文本建模等技术领域。具体实现方案为:构建源语言-目标语言平行语料数据集,通过无监督词对齐算法对平行语句构建词对齐矩阵;在Transformer结构的跨语言预训练模型的每个子层之间插入词对齐适配器,通过掩码语言建模与词对齐建模联合训练,实现跨语言表示特征的语义对齐;将所述词对齐适配器模块生成的跨语言表示特征输入至任务适配器,从而实现多种跨语言下游任务。解决了现有技术中低资源小语种难以形成词对齐映射导致的跨语言表示效果较差的问题。根据本申请的技术提高了面向低资源小语种的跨语言文本表示及多种跨语言下游任务的性能。

申请日期 2022-12-22

申请号 CN202211657931.X

公开(公告)号 115774998A

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/365002/9241>

专题 多语种信息技术研究室

推荐引用方式 杨雅婷,艾合塔木江·艾合麦提,王磊,等. 融合词对齐适配器模块的跨语言文本表示方法、装置、设备及存储介质. GB/T 7714 115774998A[P]. 2023-03-10.

**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
- 📖 [杨雅婷]的文章
- 📖 [艾合塔木江·艾合麦提]的文章
- 📖 [王磊]的文章

**百度学术**

- 📖 百度学术中相似的文章
- 📖 [杨雅婷]的文章
- 📖 [艾合塔木江·艾合麦提]的文章
- 📖 [王磊]的文章

**必应学术**

- 📖 必应学术中相似的文章
- 📖 [杨雅婷]的文章
- 📖 [艾合塔木江·艾合麦提]的文章
- 📖 [王磊]的文章

**相关权益政策**

暂无数据

**收藏/分享**

- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
-

☰ 条目包含的文件

条目无相关文件。

全部评论 (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。

QQ客服

官方微博

反馈留言