

DDiNER: 面向复杂工业场景的领域词典引导 中文命名实体识别方法

刘荣辉^{1,2}, 崔巍^{1,2}, 梁骁俊², 桂卫华^{2,3}

¹华南理工大学未来技术学院, 中国广州市, 510641

²鹏城实验室, 中国深圳市, 518055

³中南大学自动化学院, 中国长沙市, 410083

摘要: 在工业流程中, 准确的中文命名实体识别 (NER) 对于信息抽取、知识图谱构建及智能决策等应用具有重要意义。然而, 实体边界模糊、语义重叠及标注数据不足等问题严重制约其性能。针对上述挑战, 本文提出一种领域词典引导的中文NER框架——DDiNER。该框架通过层次化词典适配器将层次化工业领域词典与双向编码表示模型进行融合, 并结合双向长短期记忆网络和条件随机场实现多层级特征融合。实验结果表明, DDiNER取得优异性能, 其平均精确率、召回率和F1值分别达到95.75%、95.73%和95.74%, 显著优于现有主流方法。在独立数据集上的验证结果证实, 该模型在识别未注册实体和长尾实体方面具有良好的鲁棒性与泛化能力。本研究为工业领域中文NER提供了一种高效且可扩展的解决方案, 在下游智能应用领域具有显著潜力。

关键词: 命名实体识别; 流程工业; 领域词典; 层次化词典适配器
<https://doi.org/10.1631/ENG.ITEE.2025.0047>