
计算机网络信息中心在知识指导的科研项目分类研究中获进展

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/15539.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

为科研项目寻找到合理的领域与方向，从而协助科学家与科研管理者进行研究方向管理，这是面临的典型问题。中国科学院计算机网络信息中心大数据部博士生肖濛完成的研究成果Expert Knowledge-Guided Length-Variant Hierarchical Label Generation for Proposal Classification，发表在国际电子电气工程师学会（IEEE）旗下会议（International Conference on Data Mining 2021）上。

随着科研项目申请迅速增长，学科划分不断细化，申请者及科研管理人员如何在具有层次结构的学科体系中选择项目申请所属的细分学科领域，是需要解决的重要问题。申请书数据是由标题、摘要、关键词、额外属性、正文等不同类别数据组成的复杂结构文本。在申请书的分类过程中，专家知识可在较为容易判断的大类学科中提供对模型的指导，并利用模型生成合适粒度的层级标签。

该研究提出了一个Encoder-Decoder架构的层次多标签分类算法（Hierarchical Multi-label Classification）。编码器部分，通过利用多层次的Transformer组件实现对申请书中文本文档内的字词、文本文档间的互动关系进行建模，从而完成多类别文本的语义抽取。解码器部分，利用自注意力机制抽取专家知识以及已生成标签的特征，并结合编码器所抽取的语义信息，依照学科体系结构逐层生成标签序列。该研究为科研项目分类场景提供了结合人工智能的解决方案，拥有丰富的应用场景。

图1.专家知识指导的层次多标签生成任务

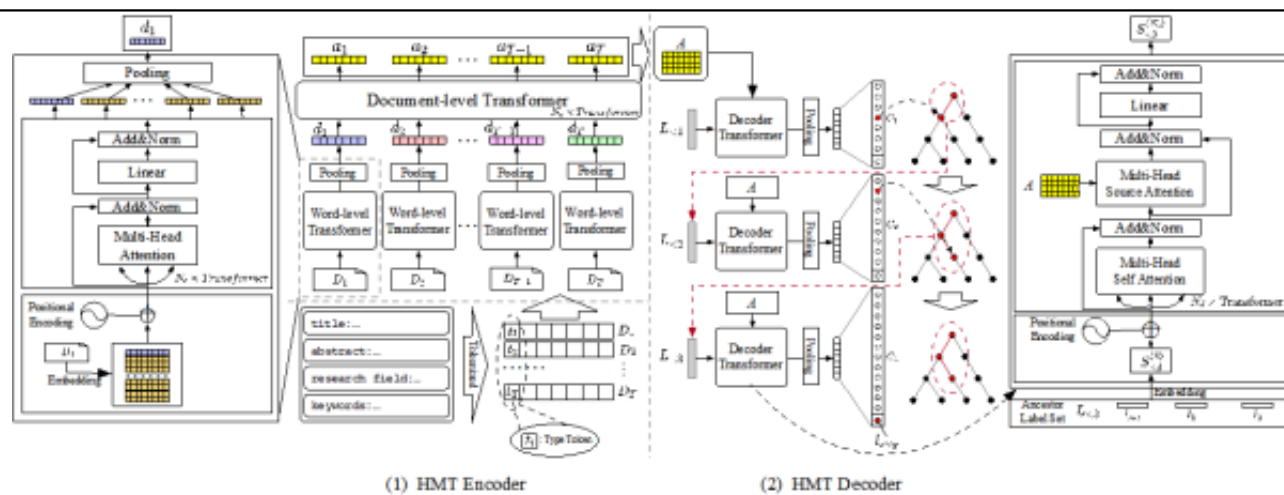


图2.专家知识指导的科研项目申请书变长层次多标签生成模型图

研究团队单位：计算机网络信息中心

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归作者所有，请勿用于商业用途，[爱科学iikx.com](http://iikx.com)转发