Automatic Classification Research of Similar Categories Based on BERT-MLDFA: Take E271 and E712.51 in CLC as an Example

LI XiangDong[1,2] SHI Jian[1] SUN QianRu[1] HE ChaoCheng[1]
(1. School of Information Management, Wuhan University, Wuhan 430072, P. R. China; 2. Center for Electronic Commerce Research and Development at Wuhan University, Wuhan 430072, P. R. China)

Abstract: This paper discusses the method of using deep learning method to improve the classification performance of the similar categories in *Chinese Library Classification* which have the features of high correlation degree and low differentiation degree. This paper proposes a BERT-MLDFA model that dynamically integrates parameters of different BERT layers through multi-level attention mechanism, and further pretrains on task datasets. Then, to conduct automated classification experiments, E271 and E712.51 in *Chinese Library Classification* were used as typical similar categories. The results show that the Macro_F1 value of the proposed method reaches 0.987, which is 2.4% higher than that of the classical machine learning method. The method proposed in this paper can capture the subtle semantic differences between texts of similar categories, which can be applied to *Chinese Library Classification* and other similar categories and is universal.

Keywords: *Chinese Library Classification*; Deep Learning; BERT; Automatic Classification

（收稿日期：2022-02-03）

征文通知

2022年中国科学技术信息研究所博士后科研工作站设站20周年学术&主题征文通知

2002年中国科学技术信息研究所获批成为国内首家具有独立招收资格的“图书情报与档案管理”学科博士后科研工作站。2022年，20周年之际，为回顾和总结图情档学科博士后科研工作站的发展历程和建设成就，中信所拟汇聚学界同人就图情档学科发展、人才队伍培养、支撑科技决策实践等内容开展征文探讨。

本次学术&主题征文均组织专家评选，设立一、二、三等奖若干，颁发荣誉证书并给予一定奖励。获奖稿件将推荐到《中国软科学》《情报学报》《情报工程》《中国科技资源导刊》《数字图书馆论坛》《全球科技经济瞭望》《高技术通讯》等期刊遴选发表。获奖作品将集结成册，收录到设站20周年纪念册。

一、学术征文选题

（1）新时代图情档学科发展研究。

（2）图情档学科博士后“学—研—产”实践创新研究。

（3）图情档学科设站单位高水平人才培养特色研究。

二、主题征文选题

（1）表达对工作站设站20周年的真情实感等。

（2）回忆或记录在中信所从事博士后科研工作的心路历程和经历故事等。

（3）介绍作为博士后合作导师的指导经验方法与心得体悟等。

（4）回顾设站20年的发展历程，以见证者、亲历者的身份记述工作站的发展变化，讲述20年来各个发展时期的重大事件、重要人物、发展成果等。

三、征文要求

（1）投稿文章必须是未公开发表的原创性研究成果。学术征文要求论据充分、数据可靠、结构严谨，字数在6 000字以上，格式参考《数字图书馆论坛》投稿模板。主题征文体裁不限，字数在2 000字以上（诗歌除外）。

（2）投稿请发送至huayf@istic.ac.cn（邮件主题注明“设站20周年学术征文/设站20周年主题征文”，论文以“文章名称—作者姓名”的方式命名，稿件内附个人简介，包括作者姓名、联系方式、工作单位、职务等基本信息）。

（3）投稿截止日期为2022年6月30日。

联系人：花老师

联系方式：010-58882046