作者简介

沈自强，男，1997年生，硕士研究生，研究方向：科技政策、自然语言处理。
李晔，男，1981年生，博士，研究员，通信作者，研究方向：深度学习、信息技术，E-mail：liye@sdas.org。
丁青艳，女，1981年生，博士，研究员，研究方向：管理系统工程、复杂系统优化。
王金颖，女，1986年生，博士，助理研究员，研究方向：科技政策、创新管理。
白全民，男，1983年生，博士，副研究员，研究方向：科技政策、创业与战略管理。

Research on Science and Technology Policy Text Classification Based on BERT Model

SHEN ZiQiang[1,2] LI Ye[1,2] DING QingYan[3] WANG JinYing[1,2] BAI QuanMin[1,2]
(1. School of Economics and Management, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, P. R. China;
2. Institute of Science and Technology for Development of Shandong, Jinan, 250014, P. R. China;
3. Shandong Computer Science Center (National Super Computer Center in Jinan), Jinan 250014, P. R. China)

Abstract: In the context of the application of smart government, this article uses deep learning methods to automatically classify massive amounts of scientific and technological policy text data in order to reduce the cost of manual processing and improve the efficiency of policy matching. This paper used the BERT deep learning model to automatically classify science and technology policies. It extracted the keywords of the policy text through the TextRank algorithm and the TF-IDF algorithm, then integrated the policy titles and policy keywords into the BERT model, so as to optimize the experiment and improve the effect and accuracy of policy text classification. It also made a comprehensive comparative analysis of the classification effect on different deep learning models to show the superiority of this method. The results show that the classification effect of combining the title and TF-IDF policy keywords is the best through the BERT model, and the accuracy rate can reach 94.41%, which proves that adding policy keywords on the basis of the title can improve the accuracy of automatic classification of policy texts on BERT model. Our research achieves an efficient classification of science and technology policy texts.

Keywords: Science and Technology Policy; Text Classification; BERT Model; Keyword Extraction

（收稿日期：2021-12-08）

书 讯

《汉语主题词表》

《汉语主题词表》自1980年问世以后，经1991年进行自然科学版修订，在我国图书情报界发挥了应有作用，曾经获得国家科学技术进步二等奖。为适应网络环境下知识组织与数据处理的需要，由中国科学技术信息研究所主持，并联合全国图书情报界相关机构，自2009年开始进行重新编制工作，拟分为工程技术卷、自然科学卷、生命科学卷、社会科学卷四大部分逐步完成。目前工程技术卷和自然科学卷已出版。

《汉语主题词表（工程技术卷）》共收录优选词19.6万条，非优选词16.4万条，等同率0.84，在体系结构、词汇术语、词间关系等方面进行了改进创新。《汉语主题词表（自然科学卷）》共收录专业术语12.4万条，包含数学、物理学、化学、天文学、测绘学、地球物理学、大气科学、地质学、海洋学、自然地理学等学科领域，收词系统、完整，语义关系丰富、严谨，每条词汇都有相应的学科分类号表现其专业属性，并与同义英文术语对应。同时，建立《汉语主题词表》网络服务系统，提供术语查询、文本主题分析、知识树辅助构建等服务。《汉语主题词表》可用于汉语文本分词、主题标引、语义关联、学科分类、知识导航和数据挖掘，是文本信息处理及检索系统开发人员不可或缺的工具。

《汉语主题词表（工程技术卷）》已于2014年由科学技术文献出版社出版，分为13个分册，总定价3 880元。《汉语主题词表（自然科学卷）》已于2018年5月由科学技术文献出版社出版，分为5个分册，总定价1 247元。两卷均可分册购买。