



效用视角下核心引文识别方法初探*

□李凌 张若楠 崔佳楠 李鑫鑫

摘要 基于引用频次开展的引文分析忽略了引文对施引文献的效用差异,依此展开资源评估、学术影响力判断难免受到“无效”引用干扰。为排除“无效”引文,提升测度及评价数据质量,文章从引文效用的角度,选择引文属性、功能、对象、情感等典型特征,建立引文标注框架。测试逻辑回归与支持向量机等方式对引文自动分类的效果,择优构建“效用视角下核心引文识别模型”。其中,为探索自动化分类方式,在数据标注阶段,除人工标引,还使用大规模预训练语言模型的对话系统 ChatGPT 进行自动分类标注,并进行建模效果测试,以期对引文自动分类方法及应用实践提供新思路。

关键词 引文效用 引文内容分析 引文标注框架 支持向量机 引文自动分类 ChatGPT

分类号 G353.1

DOI 10.16603/j.issn1002-1027.2023.06.014

引文分析是一种通过对学术文献之间的引用关系进行分析,进而对研究成果的价值做出判断的科学评价方法,其基本原理是通过分析文献被引用的次数和引用关系,揭示学术文献之间的联系和影响力^[1]。学术界也将引文分析结果作为文献或学者学术影响力、资源绩效评估的重要参考依据之一。传统引文分析主要基于引用频次来量化分析文献间的引用关系,或将引用频次作为评判资源质量、保障率、甚至学术影响力的重要标准,虽简单直观,却无法揭示文献间深层次引用关系,忽略了引文在研究中发挥的实际效用^[2]。

从内容效用的角度,引文对施引文献研究论点、研究结果的创新性、科学性、可靠性的论证和支持作用,及对论据充分性、论证合理性具有直接贡献,可充分体现出引文的必要性及其引用价值^[3]。因此,在论文中所引用的起到关键性学术论证作用、对施引文献研究的内容、方法、观点发挥了重要支撑或参考作用的引文,可视为“核心效用引文”。通过引文效用分析可尽量降低“低效”甚至“无效”的引文干扰,从而为深入开展更加科学、客观的测度和评价工

作奠定基础,对科学计量学和科学学的发展大有裨益^[4]。文章以学科期刊论文的参考文献作为研究对象,从引文效用的视角,选择引文属性、引文功能、引用对象、引用情感等特征,建立引文标注框架;并依此进行人工标注及大规模预训练语言模型的对话系统(Chat Generative Pretrained Transformer, ChatGPT)自动分类标注;测试逻辑回归分析、支持向量机等方法对引文效用分类的效果,择优构建“效用视角下核心引文识别模型”,并探讨该方法应用拓展的可行性。

1 相关研究

传统引文分析法主要基于引用频次来量化分析文献间的引用关系,但为深入揭示文献间深层次引用关系,学者们将引文分析逐步拓展至内容层面,通过对引用位置、引用频次、引用深度、引用对象等方面的分析,揭示引文之间语义关联或引用行为特征,也证实了引用效用的差异性^[5]。同时,结合不同角度对引文分类的体系、方法及相关应用的研究也在不断推进。

* 高校图书馆数字资源采购联盟(DRAA)课题“效用视角下核心引文识别与数字资源评估”(课题编号:2023DRAA03)、“西安交通大学基本科研业务费专项科研项目(人文社科类)“西安交通大学医学部学科及人才竞争情报分析研究”(课题编号:SK2018012)的研究成果之一。

通讯作者:李凌,邮箱:liling1028@xjtu.edu.cn。



1.1 引文分类体系研究

为深入揭示文献间的引用关系或引用行为,国内外很多研究分别从“引用动机”“引用目的”“引文功能”等角度对其进行分类并构建相应的引文分类体系^[5]。早期,尤金·加菲尔德(Eugene Garfield)^[6]提出基于15种引用动机构建引文索引,包括支撑、批判、评价等,为后续引文分类研究奠定了基础。随后几十年发展过程中,引文分类体系内容逐步扩充及深入。芬尼(Finney)^[7]与博尼茨(Bonzis)^[8]分别将引文位置、引用强度等特征引入了引文分类体系。奥彭海姆(Oppenheim)^[9]将引文功能细分为“历史背景”“相关工作的描述”“提供信息或数据”“比较”等7种类别。二十一世纪后,引文分类体系在指标的深度与广度方面进一步延伸,也不断推动着引文内容标注框架的发展。西蒙·泰弗(Simone Teufel)^[10]提出“议论文式的分区”引文分类法,将引文区域进一步细分为结论、证据、方法、背景、目标、未来工作等,并将该分类法应用于自动化文本信息提取。国内也有学者提出“要结合引文分布、引文次数及被引内容”综合判断引文效用^[11],并从语法和语义方面对引文属性、被引属性及两者间属性展开深入分析,构建相对全面的引文内容分析框架^[12],还出现了包含引文分类标注体系、引用对象标注体系及引文属性标注体系的面向引用关系的引文内容标注框架^[13-14],较为全面地揭示出文献引用关系及引文客观特征。

1.2 引文标注及分类方法研究

引文标注是进行合理分类的前提,早期研究均采用人工标注,虽准确率高,但难以兼顾样本量和标注效率。随着计算语言学、自然语言处理的成熟,从大规模科学文本中自动化抽取引文内容进行语义分析成为可能。随之涌现出基于规则、基于机器学习、基于深度学习、基于语义相似度的引文自动分类方法。

基于规则的引文自动分类方法通常使用手工设计的规则集进行引文分类^[15]。其优势在于专人设计和维护规则集,能提供高度的可解释性和可控性,且不需要大量的标注数据。但规则集设计受限于编辑者的知识水平,在多种复杂的关系和模式时,规则集的设计不仅难度大幅增加,可能会出现疏漏和错误,需采用有效的设计和维护策略,以确保规则集的可靠性和适应性。因此,在现实应用中,基于规则的

引文自动分类方法通常需与其他自动分类方法结合使用,以提高分类准确率和可靠性^[16]。

机器学习是引文自动分类领域中最常用的技术之一。通过使用分类器模型和特征工程等技术,研究人员可以自动将引文分为不同的类别。其中,支持向量机(Support Vector Machine, SVM)、决策树(Decision Tree, DT)、朴素贝叶斯(Naive Bayes, NB)和随机森林(Random Forest, RF)等算法得到了广泛应用^[17-18]。一些研究还探索了使用多个分类器模型进行集成的方法,以进一步提高分类的准确率^[19]。

深度学习是引文自动分类领域中的新兴技术,可自动地从数据中提取特征。其中卷积神经网络(Convolutional Neural Network, CNN)^[20]、循环神经网络(Recurrent Neural Network, RNN)^[21]等深度学习模型得到了广泛应用。近年来,使用预训练模型的方法也成为了引文自动分类的热门研究方向。但基于机器学习和深度学习的分类效果依赖于前期语料库质量及规模,训练数据质量差或语料库规模小,均会导致训练模型效果不稳定,分类优势无法发挥^[5]。

1.3 引文分类应用研究

引文分类应用主要体现在引用规律及行为揭示、引文识别、学术影响力评价等方面。如采用情感词匹配法识别论文引文的引用情感,并发现其引用行为特点及变化^[22],或采用基于词袋特征的SVM分类器对引文进行情感分类^[23]。还有通过人工标注中文图书在施引文献中引文位置、情感并计算其长度和强度,发现不同学科领域对中文图书的引用行为差异^[24]。在学术影响力评价应用中,相关研究通过对引用主题、功能和情感进行手工识别与分类,定性分析其学术影响力^[25],或基于引用情感等特征进行引文分类后加权计算单篇论文的学术影响力^[23],还有研究加入引用强度、引用位置、引用情感等因素,利用层次分析法确定引用强度、引用位置的权重,结合引用强度计算作者学术影响力AAI指标^[26]。

综上所述,学术界不断探索引文分类体系与方法,发现引文行为特征、引文规律,并以此为基础开展引文评价、学术成果影响力、学者影响力评估。但依然存在以下问题:分类体系缺乏统一标准、通用性不高,实践中仍以引用频次为主要依据,基于引文效



用分析的应用场景较少。再者,现有应用类研究中多采用先人工标引后引文分类的方法,虽已有利用深度学习进行自动分类的尝试,但由于深度学习模型更依赖于语料的规模和质量,鉴于目前训练数据偏少,其模型的优势无法发挥,效果还有待于进一步验证。

2 问题提出与方法设计

2.1 拟解决的关键问题

问题一:引文效用分类特征选择与自动提取。已有研究中提出的多样化引文特征包括:引文的年份、内容、上下文语句、频次/强度、目的、动机、功能、位置、情感等。对引文以上特征展开全面细致的标注无疑会增加标注难度,对标注者学科专业知识与信息专业技能要求较高,引文标注的准确性与特征提取的高效性很难兼顾。

问题二:效用视角下核心引文识别模型构建。在选取引文分类关键特征的前提下,构建高质量分类模型是实现引文效用精准识别的关键。现有研究中虽已有基于引文功能和情感、利用传统机器学习方法或深度学习方法进行引文自动分类,但其目的在于发现引文功能和情感的相关性,深入揭示科学发展规律、发现知识传播路径^[5],缺少对引文效用的关注及其价值的利用。

2.2 方法设计

研究以引文效用分类与核心文献提取为目标,利用相关性矩阵和散点图矩阵选择关键特征构建分类模型。再尝试利用大规模预训练语言模型、模式匹配等技术实现关键特征的自动提取。在此基础上,利用逻辑回归分析、支持向量机的方式,分别测试其对引文效用的分类效果,择优构建“效用视角下核心引文识别模型”,进而提升“核心引文”的识别效率,探索排除“低效”甚至“无效”的引文干扰的方法,为进一步的测度和评价工作奠定科学基础。研究主要分为数据获取、引文内容标注框架设计及标注、关键特征选取、分类模型建立、效果对比与分析等步骤。

在方法准确性方面,研究方法聚焦于模型对引文效用的判断与自动分类,模型的准确性由训练数据质量及模型质量决定。一方面训练数据采用人工标引方式,馆员判断引文特征,专业师生判断引文效用,充分发挥人员专业特长,最大限度确保训练数据

的详细与准确,进而保证模型效用判断的准确。另一方面,模型质量通过科恩卡帕系数及正确率评价,可判断模型是否具备良好的引文效用预测功能。

在数据保障方面,利用西安交通大学机构知识库,选取医学部 2022—2023 年部分高质量(ESI 收录、Q1 等)论文的引文作为研究对象。考虑到综述性文章的引文效用相似性较高,故去掉“综述(Review)”类型文章,选择“文章(Article)”的引文,经筛选,最终选择 1778 篇引文作为此次研究数据,并将原始数据分为训练集、测试集和预测集三部分。其中随机选取 1424 篇用作模型构建,其中 75% 的数据(1068 篇)用于识别核心引文关键特征的发现以及模型训练,25% 的数据(356 篇)用于模型效果测试。最后,选择 354 篇引文用于核心引文的识别预测。

3 引文效用分类原则

3.1 引文效用分类标准

“核心”引文在不同应用背景下具有不同内涵。一指被 SCI、EI 等权威索引收录可称之为核心;另有从文献利用的角度指被使用次数较多的高需求文献。而文章中“核心引文”指在论文中所引用的起到关键性学术论证作用,对施引文献研究的内容、方法、观点发挥了重要支撑或参考作用的引文。其重要程度需依据专业知识与信息素养综合判断,并通过五级分类法表示,即“非常重要”“重要”“一般”“不重要”“非常不重要”(具体描述见表 1)。

表 1 引文效用分类标准

引文效用	定义与描述	二分类处理
非常重要	引文作为施引文献的主要研究方法、数据及理论来源,是施引文献研究过程中不可或缺的组成部分	核心引文
重要	施引文献利用了引文中的数据、结论、方法等,如缺失将可能影响相关论点及结论	
一般	引文与施引文献中结论、方法、结果等方面起到支撑或对比作用,删除可能导致论据单薄,但并不影响研究过程与结论	非核心引文
不重要	引文作为施引文献的研究背景、相关研究、未来启示等	
非常不重要	引文对于施引文献作用不明显或不明确,去或留并无影响	



在模型构建阶段,需将引文效用作为目标值,进行二分类处理,故需对定义进行调整,其中非常重要、重要、不重要、非常不重要的分类倾向较为明确。“一般”通常为引文效用表现不突出,且容易出现效用的争议,故归为“非核心引文”。

3.2 引文内容标注框架设计及标注

本研究从引文效用角度出发,设计了包括引文特征属性、引文功能属性、引用对象属性、引用情感属性在内的引文内容标注体系(如表2所示)。引文特征属性包括:引文年份、引文所在期刊、引文出现章

节名称、引文出现频次、引文所在语句。引文功能划分为“基于、使用、支撑、比较、相关研究、未来启发、背景介绍”。引用对象是指施引文献引用引文的对象,由于医学类论文引用对象相较社科类引用对象更为简单且形式固定,故本研究将引用对象归纳为“概念理论、方法、数据、其他”。引用情感分为积极、消极、中立三种情感。由于医学论文多在于客观论证或描述,因此,对施引文献起正面支撑作用的引文均属于积极,与研究结果相悖、指出现有成果不足、研究局限等引文表示消极,其余则均标注为中立。

表2 引文内容标注框架

特征分类	引文特征	特征定义与描述
引文特征属性	引文年份	引文发表的年份
	引文所在期刊	引文所在的期刊
	引文所在章节名称	引文在施引文献中出现的章节名称,如背景介绍、方法、结果、结论、讨论等
	引文出现频次	引文在施引文献出现的频次
	引文所在语句	引文在施引文献中的语句
引文功能属性	基于	施引文献研究以引文为起点、受到启发或基于引文开展的深入研究
	使用	施引文献研究使用了引文研究的方法、成果、实验数据、结果、研究结论等
	支撑	引文的结论、数据、研究内容对施引文献研究或结论的形成起到支撑作用
	比较	施引文献研究与引文研究在结论、方法、结果等方面形成对比
	相关研究	与施引文献研究相关的研究方法、结论、数据、案例等介绍
	未来启发	引文对施引文献进一步研究有所启发,或在未来研究中可能有所涉及
	背景介绍	与施引文献工作相关的时代背景、研究背景、学术背景、文化背景等
引用对象属性	概念理论	引用被引文献中提出的概念、定义、介绍等
	方法	使用被引文献中的研究或试验方法、模型、算法、工具、方案等
	数据	使用被引文献中提出的研究数据、研究/推导公式、研究结论、应用成果等
	其他	未提及明显的引用对象
引用情感属性	积极	引文内容、主要观点或结论对施引文献的论证方法、观点、结论起支撑作用
	中立	除表现出“积极”“消极”的其他论述
	消极	引文内容、主要观点或结论与施引文献研究结果相悖、引文指出现有成果的不足、研究局限

由于引文特征判断既需要文献所涉及的学科专业知识,也需要信息专业技能,因此,选择馆员与专业师生两组人员对引文的不同类特征分别标注,有助于提升标注结果的准确性。由不同专业背景的学科馆员依据引文标注框架,分别对原始数据所有引文特征属性、功能属性、引用对象、引用情感等属性进行标引。医学专业教师及研究生若干,依据引文效用分类标准(表1)对训练集与测试集引文仅进行引文效用判断及标注。

4 自动分类模型设计与效果对比

引文特征众多,但并非所有特征都会对核心引文的判断产生影响。笔者发现,在众多引文特征中,与“引文效用”相关性较强的部分特征作为模型建立的基础尤为关键。在此基础上,进一步利用机器学习的方法构建核心引文识别模型,既可以一定程度简化引文特征标注体系,也可以提升核心引文识别效率。研究通过 KNIME 软件进行逻辑回归及支持向量机(SVM)模型的设计及训练,包括:数据读取



及观察、特征筛选、模型训练、模型验证及评价 4 个模块(如图 1 所示)。

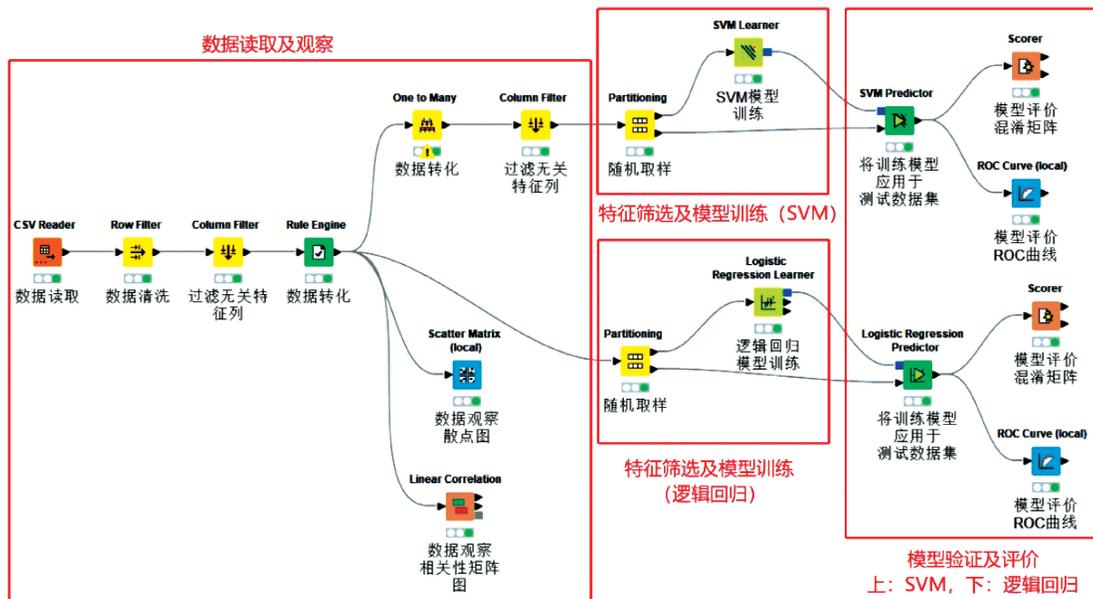


图 1 逻辑回归与支持向量机模型训练流程图

4.1 特征筛选

引文标注体系包括引文题目、引文期刊、引文年份、引文所在章节名称、引文出现频次、引用情感、引文功能、引用对象类型、引文是否重要等 9 种引文特征指标,其中引文是否重要为目标指标,通过相关性矩阵(如图 2 所示)和散点图矩阵(如图 3 所示),观察各指标和目标间的相关性。在相关性矩阵中,引文效用通过“是否重要”表示,依据表 1 中的五级分类法分为非常重要、重要、一般、不重要、非常不重要。在二分类处理后,其中非常重要、重要归为“核心引文”,一般、不重要、非常不重要归为“非核心引文”。

	引文所在章节名称	引文出现频次	引用情感	引文功能	引用对象类型	是否重要
引文所在章节名称	■ corr = +1	×	×	×	×	×
引文出现频次	×	■ corr = +1	×	×	×	×
引用情感	×	×	■ corr = +1	×	×	×
引文功能	×	×	×	■ corr = +1	×	×
引用对象类型	×	×	×	×	■ corr = +1	×
是否重要	×	×	×	×	×	■ corr = +1

图 2 指标相关性矩阵图

从相关性矩阵发现(如图 2 与图 4 所示)引文出现频次和引文效用几乎没有相关性,引文所在章节名称、引用情感、引文功能、引用对象类型等特征与引文效用相关性较强,其中引文功能相关性最高,故以此特征值为起点,依次加入其他特征值进行测试,最终选定的特征为:引文所在章节名称,引用情感,引文功能,引用对象类型。

4.2 逻辑回归模型训练及验证评价

逻辑回归(Logistic Regression, LR)算法是一种广义的线性回归分析模型,用于解决二分类问题的机器学习方法,可预测某种事物的可能性^[27]。在本研究中即通过 LR 算法训练出的模型对某引文是否为核心引文进行预测,并选用混淆矩阵(Scorer 节点),预测准确率(Accuracy)、科恩卡帕系数(Cohen's Kappa)值以及受试者工作特征曲线(Receiver Operating Characteristic Curve, ROC)^[28]进行评价。其中,科恩卡帕系数是可用于衡量分类效果的指标,即模型预测结果和实际分类结果是否一致,值越高表示该模型的分类型与实际分类结果一致性越高。预测准确率代表模型分类正确的样本数除以所有样本数,准确率越高分类器越好。训练过程中选用梯度下降算法(Stochastic Average



Gradient),设置学习率为 0.001,最终训练的模型应用到测试数据集上的结果如图 5 及图 6 所示。

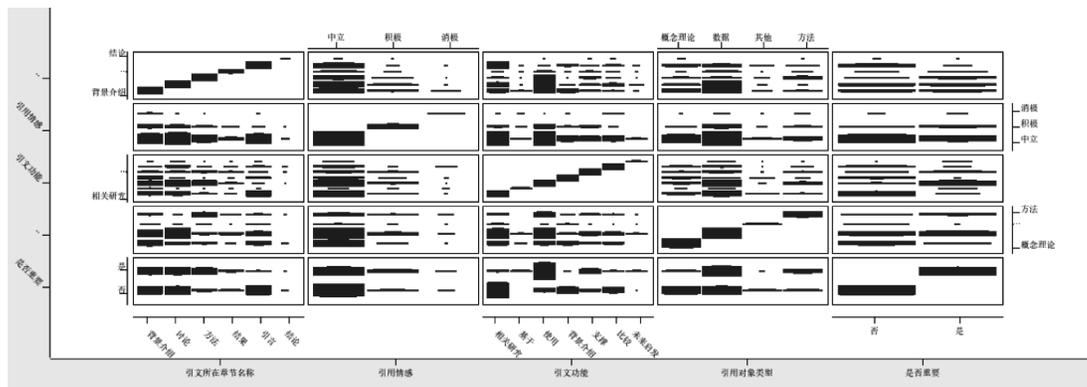


图 3 各指标散点图

Row ID	D 引文...	D 引文...	D 引用情感	D 引用功能	D 引用对象类型	D 是否重要
是否重要	0.4082830...	?	0.2492376699656...	0.722071434700...	0.426052628760...	1.0
引用功能	0.2665492...	?	0.2562750273182761	1.0	0.244650071567...	0.7220714347001006
引用对象类型	0.3061776...	?	0.2686970836013...	0.244650071567...	1.0	0.4260526287607...
引文所在章...	1.0	?	0.1897132239300...	0.266549213391...	0.306177628361...	0.4082830573885...
引用情感	0.1897132...	?	1.0	0.256275027318...	0.268697083601...	0.2492376699656...
引文出现频次	?	1.0	?	?	?	?

图 4 指标相关性矩阵参数图

是否重要 \ Prediction (是否重要)	否	是
否	167	18
是	14	157

Correct classified: 324 Wrong classified: 32
 Accuracy: 91.011% Error: 8.989%
 Cohen's kappa (K): 0.821

图 5 逻辑回归预测结果

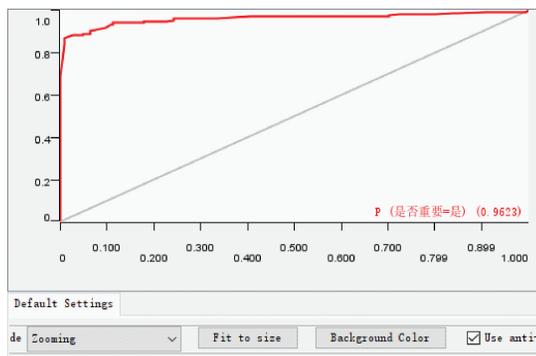


图 6 逻辑回归 ROC 曲线

训练结果发现,逻辑回归训练模型对核心引文预测结果与人工判断结果相似度已达到较高水平,在 356 条数据中,误将核心判定为非核心 14 篇,误将非核心判定为核心 18 篇,预测准确率为 91.011%,科恩卡帕系数值达到 0.82,提示该预测模型与人工判断的结果具有较高一致性。从受试者 ROC 曲线来看,曲线下方面积(Area Under Curve,AUC)很大,说明此模型预测准确率较高。

4.3 支持向量机模型训练及对比

支持向量机(SVM)是按监督学习方式对数据进

行二元分类的广义线性分类器,通常用来进行模式识别、分类以及回归分析^[29],在本研究中同样通过该算法训练出的模型对某引文是否为核心引文进行预测。SVM 算法与 LR 算法原理不同,故将两种算法训练出的模型进行对比后,择优选择。SVM 算法进行训练的流程和 LR 基本一致,但由于该算法需要所有的特征是数值,所以针对类型特征,要通过独热编码(One-hot Encoder)进行数据转化处理。在使用相同训练数据集以及测试数据集的情况下,LR 算法和 SVM 算法的模型评价结果对比如图 7 所示。



是否重要 \ Prediction (是否重要)	否	是
否	167	18
是	14	157

Correct classified: 324 Wrong classified: 32
Accuracy: 91.011% Error: 8.989%
Cohen's kappa (K): 0.821

逻辑回归(LR)

是否重要 \ Prediction (是否重要)	否	是
否	179	6
是	21	150

Correct classified: 329 Wrong classified: 27
Accuracy: 92.416% Error: 7.584%
Cohen's kappa (K): 0.848

支持向量机(SVM)

图7 LR与SVM预测结果对比图

从SVM算法训练结果可见,该算法对核心引文预测结果较LR更高,在356条数据中,该算法误将核心判定为非核心21篇,误将非核心判定为核心6篇,预测准确率为92.416%,科恩卡帕系数值为0.848。

对比可见,两个模型对“核心引文”预测准确率都很高,其中SVM算法在准确率以及科恩卡帕系数数值上表现略好。故本研究选择使用SVM算法建立的模型对预测集进行预测,并根据预测的结果进行后续研究分析。

4.4 ChatGPT自动提取引文特征及模型分类效果对比

此次建模所需的4项关键指标中,除引文所属章节名称可通过读取全文后进行模式匹配获取,引用情感、引文功能、引用对象类型均属自然语言处理(Natural Language Processing, NLP)范畴,其中引用情感主要是情感分析,引文功能可理解为对引用意图识别分类,引用对象类型则属于文本分类问题。这些均可通过深度神经网络进行学习并自动识别^[5]。2022年底,随着ChatGPT的“出圈”,ChatGPT在自然语言处理任务中文本分类、语义角色标注、机器阅读理解等方面的优势^[30]与引文内容特征分析、提取、标注与分类的需求较为契合。因此,本研究进一步测试了利用ChatGPT技术完成指

标自动获取及标注的效果。

4.4.1 特征指标自动提取及问题设计

研究始于2023年2月,使用ChatGPT 3.5版本。测试ChatGPT能否直接根据引文内容对其效用进行分类,发现效果并不理想。后对整个提取流程进行分解,先将施引文献作为背景信息,再设计特征分类提示词(Prompt),然后利用ChatGPT识别引文特征,最后写入表格文件Excel,完成特征自动提取。

首先,将施引文献作为背景信息通过共享pdf链接的方式输入ChatGPT,以确保ChatGPT完全基于施引文献背景输出特征分类。

再进行Prompt设置。Prompt可理解为分类标准的“提示语”,指输入的文本段落或短语,作为生成模型输出的起点或引导,可以是一个问题、一段描述或任何形式的文本输入^[31]。Prompt需针对“引用情感、引文功能、引用对象类型”属性及分类特征分别设置。分类依据引文内容标注框架,但为方便机器识别,将框架中引文功能简化为研究基础(基于、使用、支撑)、相关研究(比较、相关研究)、背景资料(背景介绍、未来启发)3类,经多次交互提问及分类输出效果测试后,确定关于“引用情感,引文功能,引用对象类型”特征的Prompt(如图8所示)。

是否重要 \ Prediction (是否重要)	否	是
否	179	6
是	21	150

Correct classified: 329 Wrong classified: 27
Accuracy: 92.416% Error: 7.584%
Cohen's kappa (K): 0.848

人工标引数据SVM评价结果

是否重要 \ Prediction (是否重要)	否	是
否	179	6
是	27	144

Correct classified: 323 Wrong classified: 33
Accuracy: 90.73% Error: 9.27%
Cohen's kappa (K): 0.813

自动获取数据SVM模型评价结果

图8 自动获取数据与人工标引数据SVM算法预测结果对比图

在数据输入阶段,输入引文功能及引用对象的特征分类提示词及引文列表即可识别相关特征,情



感特征提取需输入情感特征分类提示词及引文所在语句。最后,对输出结果进行格式化处理,分别输入到 Excel 文档,完成特征自动提取。

4.4.2 模型分类效果对比

为检测自动特征提取的分类效果,使用 SVM 算法及相同的训练集和测试集,分别对自动化处理后的数据和人工标记数据进行模型训练,对比发现,虽然 ChatGPT 对于引文特征的自动分类无法达到人工标引的高准确率,但在保证较好的预测效果的同时,可有效提升特征提取效率。

5 研究结果与讨论

5.1 特定应用场景下,引文效用分析比频次分析更加合理

研究核心在于排除仅依据频次无法排除的“无效”引文。故针对研究全数据集,将基于效用与基于频次的引文分析结果进行对比发现,集合 1 和 2 共有 212 种期刊重复(如图 9 所示),说明区域 2 的 212 种期刊不仅引用频次靠前,也属于核心效用引文期刊。其中,集合 1(实线部分)为利用本模型识别出核心效用引文分布的 315 种期刊;集合 2(虚线部分)为 1778 篇引文分布的 570 种期刊按引用频次降序排列的前 315 个期刊。区域 1 的 103 种期刊虽然频次较高,但从内容效用的角度并非核心效用引文,不排除存在无效甚至恶意引用的干扰。区域 3 的 103 种期刊上的引文虽引用频次不高,但却对施引文献发挥了相对较高的学术效用。由此可见,基于效用与基于频次的引文分析结果存在一定差异,单纯基于频次的引文分析无法排除无效引文干扰,也存在丢失部分有效引文的风险。基于效用的引文分析可排除部分非核心效用引文,更客观地反映出“有效”引用,而依此开展的各项应用也将更加科学。

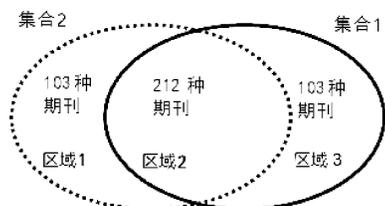


图 9 高引用频次期刊与核心引文期刊对比结果

在特定场景下,如资源保障评估、学术影响力评价等,有效引文才是评价结果科学客观的基础。行业中“他引”标准的提出与严控也正是引文效用价值

的体现。由此,单纯的频次分析做出的判断必然有失偏颇。同样,在学术热点追踪、合作网络发现过程中,过多的无效、虚假引用也将成为干扰导致“迷航”。因此在这类应用场景下,引文效用所体现出的学术论证作用尤为重要,以此为基础的数据分析也更加科学。

5.2 “ChatGPT+SVM”为高效引文特征识别与自动分类提供新思路

研究中通过 LR 方法与 SVM 训练结果对比发现,两个模型对核心引文预测的准确率都很高,其中 SVM 预测效果相对更好。对比人工与自动分类标引,自动分类虽需经过输入背景文献、提示语、相关文本,导出并格式化输出结果等步骤,但完成效率较人工逐条判断仍有大幅提升。其预测结果准确率虽略低于人工,但准确率及科恩卡帕系数值也仍处于较为理想的范围,可见“ChatGPT+SVM”模式可在保证较好预测效果的同时,显著提升特征识别与分类效率,这为批量引文特征自动提取与分类提供了新思路。

5.3 ChatGPT 的发展将为大规模引文自动分类模型提供可能

本研究针对医学学科论文使用 ChatGPT3.5 版本进行引文特征的自动提取及分类,虽然数据量及学科范围有一定局限,但其提取及分类效果已初现端倪。研究中采用少样本学习(Few-shot Learning)模式,需反复给出提示词进行预训练。而问答字数限制,也使问答过程中出现“记忆缺失”的现象,影响分类效率。

2023 年 3 月,随着 GPT-4 的发布,ChatGPT 对话生成能力、语言理解能力、模型可定制性等有所提升。于本研究而言,一方面,要客观看待 ChatGPT 对于引文自动分类的作用。引文效用分类首要依据便是施引文献内容,其次仍需相关专业及信息知识。ChatGPT 在语料库增加、对话能力增强等方面的进步,提升了其对文献的理解能力,但对文献内容并无影响。换言之,在施引文献内容、引文相关属性、分类原则确定的前提下,引文效用具备一定的客观性,不会发生较大偏差。另一方面,ChatGPT 所具备的微调(Fine-tuning)功能为构建大规模引文自动分类模型提供了可能。微调功能相当于利用在大数据集上预训练好的模型,通过自有的数据进行模型微调,即在底层数据上进行模型的“特殊训练”。因此,可



根据引文效用分类的具体应用场景,准备一定量训练数据,设定引文分类标准,实现引文分类定制化模型微调模型(Fine-tuned Model),既降低了构建超大规模模型训练语料库的操作难度,又可得到比直接用自己数据训练更好的效果。

6 结语

本文从引文效用的角度构建核心引文识别模型并探索了利用 ChatGPT 实现过程自动化的可行性。为基于引文内容分析、引文自动分类及应用提供了一种新思路。但本研究尚在探索阶段,一方面数据选取存在学科与数据量的局限性,可能导致学科资源呈现不全面。另一方面,利用 ChatGPT 3.5 版本尚未充分挖掘并呈现出其对于引文特征识别与分类的显著优势,对引文特征识别与提取的自动化程度也有待深入探索。在后续研究中,将扩大取样范围及数量,也会进一步探索基于各种大型语言模型提高引文自动化分类的效率与质量的方法与路径。

参考文献

- 1 周周杰. 引文分析的困局与科学评价的转向[N]. 中国社会科学报, 2023-08-04(004).
- 2 周志超. 国内外引文内容分析研究进展[J]. 情报杂志, 2021, 40(12): 177-185.
- 3 朱大明. 研究型论文中“关键引文”概念初探[J]. 中国科技期刊研究, 2015, 26(11): 1161-1165.
- 4 赵蓉英, 曾宪琴, 陈必坤. 全文本引文分析——引文分析的新发展[J]. 图书情报工作, 2014, 58(9): 129-135.
- 5 欧石燕, 凌洪飞. 引用文本自动分类及其应用研究[J]. 图书情报工作, 2022, 66(16): 125-136.
- 6 Garfield E. Can citation indexing be automated? [C]// Symposium on Statistical Assoc Methods for Mechanized Documentation, 1964: 189-192.
- 7 Finney B. The reference characteristics of scientific texts[D]. City University (London, England), 1979.
- 8 Bonzi S. Characteristics of a literature as predictors of relatedness between cited and citing works[J]. Journal of the American Society for Information Science, 1982, 33(4): 208-216.
- 9 Oppenheim C, Renn S P. Highly cited old papers and the reasons why they continue to be cited[J]. Journal of the American Society for Information Science, 1978, 29(5): 225-231.
- 10 Teufel S, Sidda R, Than A. Automatic classification of citation function[C]// Proceedings of the 2006 conference on empirical methods in natural language processing. Stroudsburg, PA: Association for Computational Linguistics, 2006: 103-110.
- 11 李燕. 关于引文分布和引文效用的调查分析[J]. 现代情报, 2010, 30(6): 145-147.
- 12 Ding Y, Liu X, Guo C, et al. The distribution of references across texts: some implications for citation analysis[J]. Journal of Informetrics, 2013, 7(3): 583-592.
- 13 陆伟, 孟睿, 刘兴帮. 面向引用关系的引文内容标注框架研究[J]. 中国图书馆学报, 2014, 40(6): 93-104.

- 14 张梦莹, 卢超, 郑茹佳, 等. 用于引文内容分析的标准化数据集构建[J]. 图书馆论坛, 2016, 36(8): 48-53.
- 15 Khoshgoftaar T, Seliya N, Gao K. Detecting noisy instances with the rule-based classification[J]. Intelligent Data Analysis, 2005, 9(4): 347-364.
- 16 Uzuner Ö, Zhang X R, Sibanda T. Machine learning and rule-based approaches to assertion classification[J]. Journal of the American Medical Informatics Association, 2009, 16(1): 109-115.
- 17 杨宁, 张志强. 基于机器学习的科学数据正式引用识别方法研究[J]. 情报杂志, 2022, 41(2): 182-189.
- 18 徐琳宏, 丁翌, 林原, 等. 基于机器学习算法的引文情感自动识别研究——以自然语言处理领域为例[J]. 现代情报, 2020, 40(1): 35-40, 48.
- 19 Néveol A, Wilbur W J, Lu Z. Extraction of data deposition statements from the literature: A method for automatically tracking research results[J]. Bioinformatics, 2011, 27(23): 3306-3312.
- 20 Bakhti K, Niu Z, Yousif A, et al. Citation function classification based on ontologies and convolutional neural networks [C]// Proceedings of the International Workshop on learning Technology for Education in Cloud. New York: Springer Nature, 2018: 105-115.
- 21 Yousif A, Niu Z, Chambua J, et al. Multi-task learning model based on recurrent convolutional neural networks for citation sentiment and purpose classification[J]. Neurocomputing, 2019, 335(3): 195-205.
- 22 Yan E, Chen Z, Li K. Authors' status and the perceived quality of their work: measuring citation sentiment change in Nobel articles[J]. Journal of the Association for Information Science and Technology, 2020, 71(3): 314-324.
- 23 耿树青, 杨建林. 基于引用情感的论文学术影响力评价方法研究[J]. 情报理论与实践, 2018, 41(12): 93-98.
- 24 章成志, 李卓, 赵梦圆, 等. 基于引文内容的中文图书被引行为研究[J]. 中国图书馆学报, 2019, 45(3): 96-109.
- 25 刘盛博, 王博, 唐德龙, 等. 基于引用内容的论文影响力研究——以诺贝尔奖获得者论文为例[J]. 图书情报工作, 2015, 59(24): 109-114.
- 26 李铮, 邓三鸿, 孔嘉, 等. 学者学术影响力识别研究——基于引文全数据的视角[J]. 图书情报工作, 2020, 64(12): 87-94.
- 27 朱元涛. TensorFlow Lite 移动端深度学习[M]. 北京: 机械工业出版社, 2022: 33-35.
- 28 Barbara W, Kubiak K B, Paulina J. Cohen's kappa coefficient as a measure to assess classification improvement following the addition of a new marker to a regression model[J]. International Journal of Environmental Research and Public Health, 2022, 19(16): 10213.
- 29 王征, 李晓波. Python 从入门到精通[M]. 北京: 中国铁道出版社, 2020: 174-175.
- 30 Min B, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: a survey [EB/OL]. [2023-03-29]. arXiv: 2111.01243. https://arxiv.org/abs/2111.012.
- 31 黄峰达. Prompt 编写模式[EB/OL]. [2023-03-29]. https://new.qq.com/rain/a/20230324A096NC00.

作者单位: 西安交通大学图书馆, 陕西西安, 710049

收稿日期: 2023年5月16日

修回日期: 2023年10月10日

(责任编辑: 支娟)

(转第 127 页)



Competency Transfer Oriented Disciplinary Information Literacy Education :

A Review of U.S. Business Information Literacy Guidelines and Its Implications

Tan Dandan Ke Ping

Abstract: Business information literacy, as a transferable generic competency, is valuable for solving information problems related to business topics in the contexts of academics, workplace, and everyday life. Achieving effective transfer of this competency in different contexts is a core goal of business information literacy education. This study discusses the framework of understanding business information literacy from the perspective of competence transfer and interprets the practical points of the latest guideline of business information literacy education, Business Research Competencies, from four aspects: information sources, learning subjects, classification of subject concepts, and learning outcomes. This study also points out its implications for business information literacy education in libraries in China in terms of practice basis, teaching objectives, teaching strategies, assessment of learning outcomes, and synergistic cooperation.

Keywords: Competency Transfer; Disciplinary Information Literacy Education; Business Information Literacy; Business Research Competencies

(接第 106 页)

A Preliminary Exploration of Core Citation Recognition Method from the Perspective of Utility

Li Ling Zhang Ruonan Cui Jianan Li Xinxin

Abstract: The citation analysis based on the citation frequency ignores the difference in the utility of citations to the cited documents, and it is inevitable that the resource evaluation and academic influence judgment based on this will be interfered by “invalid” citations. To eliminate “invalid” citations and improve the quality of measurement and evaluation data, the article selects typical features such as citation attributes, functions, objects, and emotions from the perspective of citation utility, and establishes a citation annotation framework, in order to test the effectiveness of logistic regression and support vector machine methods on automatic citation classification in annotated datasets, and to construct a “core citation recognition model from the perspective of utility”. Specifically, in the data annotation stage, in addition to manual indexing, a large-scale pre trained language model dialogue system ChatGPT is also used for automatic classification and modeling performance testing, in order to provide new ideas for automatic citation classification methods and application practices.

Keywords: Citation Utility; Content Analysis of Citations; Citation Annotation Framework; Support Vector Machine; Automatic Citation Classification; ChatGPT