

基于 BP 神经网络的学术论文评价模型研究

韩 雷

(浙江理工大学经济管理学院, 浙江 杭州 310018)

摘 要: [目的/意义] 学术论文是学者科研水平与学术贡献的重要佐证和标志。构建科学的学术论文评价模型, 对人才评价、科研经费分配、评奖评优、晋升及招聘等都具有重要指导意义。[方法/过程] 文章选择 Web of Science 数据库中“Information Science and Library Science”学科类别下发表于 2010 年的论文作为研究对象。首先, 基于论文多方面的关联特征构建模型特征空间; 然后, 采用机器学习中被广泛用于预测任务的有监督学习算法——BP 神经网络训练模型, 并进行十折交叉验证确保模型稳定性; 最后, 通过计算模型的校正决定系数($R^2_{adjusted}$)和均方根误差(RMSE), 选择出最优模型。[结果/结论] 本研究构建的最优 BP 神经网络模型的校正决定系数($R^2_{adjusted}$)达 0.91, 均方根误差(RMSE)约 19.8, 评价性能较好。

关键词: 学术论文; 评价模型; BP 神经网络; 科学价值

DOI: 10.3969/j.issn.1008-0821.2024.02.015

[中图分类号] G250.252 [文献标识码] A [文章编号] 1008-0821 (2024) 02-0170-08

Research on Academic Paper Evaluation Model Based on BP Neural Network

Han Lei

(School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: [Purpose/Significance] Academic papers serve as crucial evidence and indicators of scholars' research level and academic contributions. Constructing a scientific academic paper evaluation model holds significant guiding implications for talent assessment, research fund allocation, awards, promotions, recruitments, and more. [Methods/Process] This study selected papers published in the “Information Science and Library Science” discipline category in the Web of Science database in the year 2010 as the research subjects. Firstly, a model feature space was constructed based on various correlated characteristics of the papers. Then, a supervised learning algorithm widely used in predictive tasks in machine learning—the BP neural network—was employed to train the model. Ten-fold cross-validation was conducted to ensure model stability. Finally, by calculating the model's adjusted coefficient of determination ($R^2_{adjusted}$) and root mean square error (RMSE), the optimal model was selected. [Results/Conclusion] The optimal BP neural network model constructed in this study achieved an adjusted coefficient of determination ($R^2_{adjusted}$) of 0.91 and an RMSE of approximately 19.8, demonstrating good evaluation performance.

Key words: academic paper; evaluation model; BP neural network; scientific value

2016 年, 中共中央印发《关于深化人才发展体制机制改革的意见》中提到, “基础研究人才以同行学术评价为主, 应用研究和技术开发人才突出市场评价, 哲学社会科学人才强调社会评价”; 之

后, 中共中央办公厅、国务院办公厅分别于 2017 年和 2018 年印发的《关于深化职称制度改革的意见》和《关于深化项目评审、人才评价、机构评价改革的意见》中均提出类似的意见。2018 年, 国务

收稿日期: 2023-10-29

作者简介: 韩雷 (1979-), 女, 讲师, 博士, 研究方向: 金融大数据挖掘, 信息计量和科学评价。

院印发的《关于优化科研管理提升科研绩效若干措施的通知》中提到,“建立以创新质量和贡献为导向的绩效评价体系,准确评价科研成果的科学价值、技术价值、经济价值、社会价值、文化价值”。同年,教育部办公厅印发的《关于开展清理“唯论文、唯帽子、唯职称、唯学历、唯奖项”专项行动的通知》中明确提出“推行代表作评价制度,注重标志性成果的质量、贡献和影响”^[1]。可见,对人才评价不仅要求“以成果的质量评价为核心”,还推行代表作评价制度,强调对标志性成果的质量、贡献和影响的衡量。

在实际应用中,如招聘、晋升、评奖等科研管理与决策期间,通常都要求提供近 3 年或 5 年内发表的代表性成果,特别是在对代表性学术论文进行评价时,累计被引次数和所属期刊影响因子常被作为重要参考依据。然而,对于最新发表的论文,在被评价时其被引次数还未得以体现;对于一些研究过于超前的开拓性研究成果,可能需要很长时间才能引起人们的广泛关注^[2];又因“以刊评文”具有很多局限性^[3]。因此,构建适用于在论文发表早期对其进行评价的模型迫在眉睫且具有很强的应用价值。

评价学术论文的方法包括基于研究内容的定性评价法和基于关联特征的定量评价法。关于定性评价,使用最广泛的方法是同行评议。但在实际执行过程中,发现同行评议存在严重的缺陷和不足^[4]。因此,从论文关联特征层面对学术论文进行定量评价仍占重要地位。评价学术论文的核心是评价其科学价值。科学价值的判断一方面来源于论文自身对尚存问题或潜在现象、规律等的解决或发现,即创新性;另一方面来源于论文对未来科学发展或进步的促进及影响作用,即影响力。因此,本研究在综述有关学术论文评价的研究时,综合关注了论文创新性评价和影响力预测等方面的研究。此外,因被引次数是论文影响力的重要体现,因此也重点阅读了被引次数预测相关方面的研究。

相关研究中涉及的论文关联特征主要可概括为 4 类。第一类为论文自身特征,包括:一是论文外表特征,如标题^[5]、关键词^[6]、参考文献^[7]、署名作者数量^[8]、作者所属国家(地区)^[9]等。二是论文内部特征,如主题受欢迎度(Topic Rank)、论

文受众的广度(Diversity)、论文发表的新近度(Recency)^[10],从标题、摘要、关键词中提取出的元数据语义特征^[11],以及从论文题目、摘要、关键词中提取出的主题词组合新颖性^[12]等。第二类为可反映期刊影响力的特征,如期刊总被引次数、影响因子、即时指数、Eigen Factor Score、5 年影响因子等^[7]。第三类为可反映作者过往科研实力的特征,如作者过去发文数、过去发文的被引次数、过去发文所属期刊的影响力、H 指数、合作规模等^[13-14]。第四类为可反映论文短期影响力的特征,如动态被引次数^[15]、Altmetrics 指标^[16]等。综述相关研究,发现影响或可用来间接衡量学术论文创新性或未来影响力的因素很多,且影响机制非常复杂。可见,在评价学术论文时仅依赖个别特征或采用基于多维特征的传统权重确定方法都具有很大的局限性。

本研究主要是为满足现实需求而开展的,因此构建的模型不仅要评价性能好,而且需具备通用性高、执行简单等特点。本研究在以下几方面进行了创新。一是在特征空间构建方面,综合已有相关研究挑选了数据易获取且涵盖范围尽可能全面的多方面特征,同时,所有特征数据均需在论文发表后两年前就可获得。二是在训练模型的算法选择方面,采用了被广泛用于预测任务且预测性能明显优于其他很多机器学习算法的 BP 神经网络^[17]。三是在训练模型的目标值选择方面,选择将论文发表 10 年后的累计被引次数作为论文科学价值的目标值或真实值。在本研究选择的图书情报学领域,目前仅有个别研究预测过论文发表 5 年后的被引次数^[7]。

1 学术论文评价模型的特征空间

本研究所构建模型特征空间如表 1 所示,主要包括 4 个方面,17 个特征。

在构建模型特征空间时主要坚持特征数据易获取且特征类型全面性兼顾的原则。同时,在选择特征时还结合了以往研究的发现。第一,在选择论文自身特征时,主要参考了相关研究的结论,即作者数量、参考文献数、作者国际合作程度等是与论文创新性或影响力相关性较强的特征。本研究首次将 WOS 学科分类作为评价学术论文的特征。另外,由于论文内部特征如论文主题受欢迎度、主题词新颖性以及元数据语义特征等的计算均非常复杂,甚至

表 1 模型特征空间
Tab. 1 Model Feature Space

序号	特 征	特 征 描 述	所属类别
x_1	作者数	论文中署名作者人数	论文特征
x_2	国际合作程度	论文署名作者所属的国家(地区)数	
x_3	参考文献数	论文引用的参考文献篇数	
x_4	跨学科程度	论文在 Web of Science 数据库中对应的 WOS 学科分类数	
x_5	影响因子 (IF)	期刊的 Impact Factor	期刊特征
x_6	总被引次数	期刊累计总被引次数	
x_7	特征因子	期刊的 Eigenfactor Scores	
x_8	载文量	当年期刊发表论文数	
x_9	发文数	第一作者过去发表的论文总数	作者特征
x_{10}	总被引次数	第一作者过去所有发表论文的被引次数总和	
x_{11}	均被引次数	第一作者过去发表论文的平均被引用次数	
x_{12}	最大被引次数	第一作者过去发表所有论文中的单篇最大被引数	
x_{13}	平均影响因子	第一作者过去发表论文所属期刊的平均影响因子	
x_{14}	最大影响因子	第一作者过去发表论文所属期刊的最大影响因子	
x_{15}	H 指数	第一作者发表样本论文前的 H 指数	
x_{16}	均合著人数	第一作者过去发表论文的平均合著人数	
x_{17}	被引次数	发表两年内的总被引次数	早期引文特征

需要采用本文挖掘或深度学习等技术才能计算获取到。因此没有被纳入特征空间中。第二,在选择可反应期刊影响力的特征时,选择在 Clarivate Analytics 提供的期刊引文报告中对于每种期刊都被填写的特征。第三,在选择可反映论文作者过往科研实力的特征时仅选择了第一作者的相关特征,这一行为主要参考了 Wang M 等^[18]的研究发现,即在论文的关联特征中,对论文未来引用影响最显著的特征是第一作者的声誉。第四,在选择可反映论文短期影响力的特征时,没有选择替代计量学指标 (Altmetrics),主要归因于赵蓉英等^[16]的研究发现,即 Altmetrics 测度的影响力偏向社会影响力及社会关注度,较少涉及表征论文质量的学术影响力。

2 数据收集、预处理与分析

2.1 数据收集

Web of Science 数据库是国际公认的可反映科学研究水准的数据库。本研究以该数据库作为数据来源数据库。首先,选择发表于 2010 年图书情报学领域所有 SSCI 期刊中的论文作为研究对象,文献类型只选择“Article”,经对数据进行筛选过滤,最后收集到的有效论文数为 2 735 篇。然后,分别

计算每篇论文对应的特征数据。其中论文特征都可以通过直接从数据库中导出论文完整记录来计算获取。期刊特征可以通过每个期刊对应 2010 年的引文报告获取到。另外,需要特别说明的是,计算论文发表两年内获得的累计被引次数时忽略了发表在 2010 年不同月对时间的影响,因此都统一计算到截至 2012 年底获得的累计被引次数。在计算作者特征时,考虑同一作者在不同论文中的署名可能是简写也可能是全称,因此在该阶段,本研究进行了严格的数据比对与核查,以确保准确计算每个作者特征。

2.2 数据预处理

本研究按照式 (1) 将每个特征值数据都归一化到 [0,1] 之间,这是归一化方法中最常见的最大最小值法,它能够很好地保存数据的原始意义。

$$x_{new} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

2.3 数据分析

2.3.1 描述性统计分析

经计算发现,本研究所收集的 2 735 篇论文截至 2020 年累计获得的总被引次数近 8 万次,其中 6.4 万的被引次数是由近 550 篇论文获得,也就是

约 80% 的累计总被引次数由 20% 的论文获得, 可以看出其引文分布整体符合二八定律, 这也说明本研究的样本数据是有效的, 可用于研究分析。另外, 就单篇论文而言, 有些论文 10 年内没有获得一次引用, 即截至 2020 年获得的总被引次数为 0 次; 也有论文 10 年内获得的累计总被引次数高达 1 826 次。经对论文特征进行统计分析, 80% 的论文的国际合作程度特征数据为 1, 均有一个国家(地区)的作者完成。30% 的论文不存在跨学科研究, 跨学科程度特征数据为 1, 另外 45% 的论文跨学科程度特征数据为 2, 即该论文的研究内容跨两个学科领域, 其他论文的跨学科程度特征数据最多为 5。发现超 30% 的论文只有 1 个作者, 一半多论文的署名作者在两个以上, 但不超过 5 个。这些论文的参考文献数大部分多于 5 篇, 但少于 50 篇。经对期刊特征进行统计, 约 50% 的论文发表在影响因子小于 1 的期刊上, 还有近一半的论文发表在影响因子大于 1 且小于 3 的期刊上, 很少有论文发表在影响因子大于 3 的期刊上。经对第一作者特征进行统计, 约 1/3 的作者是第一次论文, 即本研究采集的该篇论文, 也就是说约 30% 的第一作者过去所有发表论文的被引次数总和为 0, 过去发表论文超过 4 篇的作者不超过 1/4; 80% 论文的第一作者过去发文所属期刊在 2010 年时的最大影响因子不足 3, 平均影响因子低于 2, 获得的累计总被引次数少于 56 次, 平均被引次数少于 14 次, 最大被引次数少于 33 次, H 指数小于 3。经对论文发表后两年内获得的被引次数进行统计, 发现累计被引次数为 0 的论文约占 1/4, 仅获得 1 次引用的论文占不足 20%。

2.3.2 相关性分析

斯皮尔曼(Spearman)相关系数是一种非参数的计算相关系数的方法, 它可以反映两个变量之间的单调关系强度。本研究为了解所构建特征空间中每个特征与目标值的相关关系, 利用统计分析软件 SPSS 分别计算了各特征之间以及每个特征与目标的 Spearman 相关系数。相关系数的绝对值越大, 相关性越强, 相关系数越接近 0, 相关度越弱。通常情况下, 通过以下取值范围判断变量的相关强度: 0.0~0.2 极弱相关或无相关; 0.2~0.4 弱相关;

0.4~0.6 中等程度相关; 0.6~0.8 强相关; 0.8~1.0 高度相关。

1) 论文特征。首先对论文特征中的论文作者数、国际合作程度、参考文献数和跨学科程度进行相关性分析, 发现这些特征之间的相关性都很弱, 论文特征中的这 4 个特征所提供的信息都是不独立、不重复的。另外, 分别计算论文作者数、国际合作程度、参考文献数和跨学科程度与目标值(即发表 10 年后获得的总被引次数)的相关性, 发现呈现中等程度相关的特征仅有 1 个, 即论文的参考文献数, 其他特征均呈现弱相关, 甚至论文的国际合作程度与目标值的相关性极弱, 也可以理解为几乎不相关。

2) 期刊特征。首先分析期刊特征中的影响因素、总被引次数、特征因子和期刊总论文数之间的相关性, 经计算发现, 除了期刊影响因素和期刊总被引次数呈中等程度相关外, 其他 4 个特征间的两两相关性都呈现高度相关。再分析各特征与目标值的相关性, 发现期刊影响因素、总被引次数和特征因子均与目标值呈中等程度的相关, 而期刊总被引次数与目标值的相关性较弱。综合来看, 期刊影响因素和期刊总被引次数是提供信息相对独立的特征。

3) 作者特征。首先分析作者特征中发文数、总被引次数、均被引次数、最大被引次数、平均影响因子、最大影响因子、H 指数和平均合著人数 8 个特征之间的相关性。经计算发现, 总被引次数、平均被引次数、最大被引次数及 H 指数 4 个与被引次数相关的特征之间的相关性很高, 这些特征中与目标值相关性最强的特征是总被引次数。除此之外, 平均影响因子和最大影响因子之间的相关性也很高, 其中平均影响因子和目标值的相关性最强。综合来看, 能提供相对独立信息且与目标值相关性较强的指标为篇均被引次数、期刊最大影响因子和平均合著人数。另外还发现, 发文数与目标值的相关性最弱。

4) 早期引文特征。早期引文特征只包括 1 个, 即论文发表两年内获得的总被引次数。经计算该特征与目标的相关系数, 发现论文发表后两年内获得的累计被引次数与目标值的相关性较强。

综上所述, 从特征所属大类来看, 短期引文特

征与论文发表 10 年后获得的总被引次数的相关性最强,其次是期刊特征、论文特征,相关性最弱的是作者特征。因本研究将采用 BP 神经网络算法对所构建的模型进行训练,为提高模型性能,需要对强线性相关的特征进行挑选,即挑选出每类特征中与目标值相关性较强且提供信息相对独立的特征,以防止各特征间多重共线性现象的出现,从而影响模型性能。经挑选,最后从 17 个特征中挑选出 9 个特征用于训练模型。挑选出的 9 个特征分别是:作者数、参考文献数、跨学科程度、期刊影响因子、期刊总论文数、第一作者过去发文的篇均被引次数、期刊最大影响因子、平均合著人数以及论文发表两年内获得的总被引次数。最后,为进一步确保挑选出来的 9 个特征之间不存在共线性现象,还对这 9 个特征进行了共线性诊断,发现这些特征之间的共线性几乎不存在,说明用这 9 个特征进行 BP 神经网络模型训练是可行的、科学的、合理的。

3 研究方法

3.1 BP 神经网络

本研究采用机器学习中一种常用的有监督学习算法——BP 神经网络,构建适合于在论文发表早期对其科学价值进行评价的模型。BP 神经网络由 Rumelhart D E 等^[19]提出,是预测任务中应用最广泛的人工神经网络模型之一。该网络的应用领域包括金融、会计、管理科学、电子商务和计算机科学。人工神经网络在文献计量学中也得到了广泛的应用。

与线性回归模型不同,BP 神经网络对数据的分布没有严格的要求。该网络的预测结果通常是鲁棒的。此外,浅层机器学习模型(如支持向量机和线性回归)的性能取决于特征功能质量。然而,对于人类专家来说,设计有用的功能并不容易。相比之下,深度神经网络在特征学习方面具有优势,即它可以通过多层非线性变换,将初始的“底层”特征表示自动转换为“高层”特征。一个完全连接的前馈神经网络可以在任何期望的精度水平上近似任何连续函数。同时,多项研究也证实了 BP 神经网络在预测方面的优势。例如,国外学者 Wong T C 等^[20]发现,BP 神经网络的性能明显优于线性回归和支持向量机;Lee S 等^[21]发现,BP 神经网络模型的预测性能优于线性回归模型;Ruan X 等^[17]

也发现,BP 神经网络的预测性能明显优于随机森林、线性回归、支持向量机、K 近邻算法和循环神经网络等。国内学者贺婉莹^[22]发现,在多元线性回归、回归树、随机森林和人工神经网络这几种机器学习算法中,神经网络的预测效果最好;胡泽文等^[23]在预测高价值专利时采用 BP 神经网络的主要原因是其性能优于卷积神经网络、长短期记忆网络、K 邻近算法和分类决策树等。因此,本文选择 BP 神经网络进行模型训练。

3.2 模型性能评估指标

决定系数(R^2)和均方根误差(RMSE)是两个流行的标准,可用来评估本研究所构建的神经网络模型的性能。首先是决定系数(R^2),其次是均方根误差(RMSE)。

R^2 衡量的是实际值和预测值之间的相关性,用于描述自变量对因变量的解释程度,可以反映模型对数据的拟合能力。 R^2 的值总是小于或等于 1,其中 $R^2=0$ 表示不相关,拟合能力差; $R^2=1$ 表示预测值与实际值完全相关,决定系数(R^2)的值越大说明训练出的模型的拟合性能越好,预测值越接近目标值。通常,决定系数(R^2)高于 0.4,就说明该模型的拟合效果较好,可以接受。另外,因研究所构建特征空间中的特征数量和收集到的样本数量影响着决定系数(R^2)的大小,特征数量越大、样本数量越多,则决定系数(R^2)的值会相应增大。因此,本研究采用校正决定系数($R_{adjusted}^2$)来代替决定系数(R^2),以规避特征数量和样本数量对决定系数(R^2)值的影响,从而提高对模型性能评估的有效性。均方根误差(RMSE)测量预测值与实际值的变化,RMSE 的值越低越好。因在计算均方根误差(RMSE)时首先要计算平均误差,而噪声点对平均误差的影响较大,如果数据中确实存在噪声点,则会导致模型对该噪声点的预测非常不准确,这将非常影响模型的平均误差值,从而影响均方根误差。因此,将均方根误差用于对模型性能的辅助评价。

R^2 、 $R_{adjusted}^2$ 和 RMSE 的计算公式分别如式 (2)~式 (4)。

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (2)$$

$$R_{adjusted}^2 = 1 - \frac{(1-R^2)(N-1)}{N-P-1} \quad (3)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

其中, N 表示样本数量, y 表示样本的实际值, $\hat{y}$ 表示样本的预测值, $\bar{y}$ 表示样本实际值的平均值, P 表示特征数量。

4 结果与分析

本研究采用信息正向传播误差反向传递的 BP 神经网络构建评价学术论文的模型。使用的是具有很强数值计算功能的 Matlab 软件中的 Newff 函数。实验时采用十折交叉验证来评估模型的性能,即将样本集分为 10 个子集,每次选择其中 1 个子集作为测试集,其余 9 个子集共同作为训练集。在设计模型时需要确定神经网络的拓扑结构和系列参数。

1) 确定网络层数及节点数

因经相关性分析后最终确定的特征数只有 9 个,对于较少的特征数通常构建只包含 1 个隐含层的网络即可满足需要。模型输入层节点数为 9,即经相关性分析后筛选出的特征个数。输出层节点为 1,即目标值。隐含层节点数确定的经验公式为 $p = \sqrt{m+n} + a$, m 为输入层节点数, n 为输出层节点数, a 是可以自取的任意值变量,通常 a 的范围可取 1~10,所以隐含层节点数的取值范围确定为 5~13。确定隐含层节点数的初始值为 5,然后由少增多依次进行实验,最终根据实验结果确定最优的节点数。

2) 选择训练函数

因对数据预处理时将数据归一化到 $[0, 1]$ 之间,因此 Log-sigmoid 型函数被应用于输入层与隐含层之间。线性传递函数——Purelin 型函数被应用于隐含层和输出层之间。

3) 选取初始权值的

本研究选择将初始权值都设置为 0。

4) 确定学习速率

学习速率太大,系统将会非常不稳定;学习速率太小,则会在保证误差范围的前提下使模型的训练时间太长。通过参考前人研究以及多次实验结果的测试,最终将学习速率确定为 0.01。

5) 停止条件的确定

模型停止训练需要满足一定的条件,通常有两种方法。一种方法是迭代次数达到设定的最大值时训练停止,还有一种方法是通过控制误差范围来作为模型训练停止的条件,只要有一种条件出现,则训练停止。通常,会有多个满足条件的网络可以被训练出,最终需要通过对结果进行分析,然后从这些网络中选择一个最优的网络。本研究选择将训练目标最小均方误差设置为 0.0001,最大迭代次数设置为 1 000。

以上 2)~5) 中参数确定后将保持不变,然后针对每组数据集分别调节对应的隐含层节点数,每设定一次隐含层节点数后都进行至少 1 000 次运行。经多次实验后,发现在隐含层节点数为 5 时,训练出的模型性能最好,因此最终确定神经网络采用 9-5-1 的拓扑结构。最后,采用训练出的 10 个神经网络模型,分别对测试集样本进行预测,并计算出每个模型对应测试集的校正决定系数 $R_{adjusted}^2$ 和均方根误差 $RMSE$,如表 2 所示。

表 2 每个神经网络模型的测试性能

Tab. 2 Testing Performance of Each Neural Network Model

模型 编号	$R_{adjusted}^2$	$RMSE$	模型 编号	$R_{adjusted}^2$	$RMSE$
1	0.58	77.1	6	0.62	38.1
2	0.65	28.2	7	0.73	24.6
3	0.83	23.4	8	0.75	29.8
4	0.91	19.8	9	0.64	31.2
5	0.55	30.1	10	0.70	26.8

计算得出 10 个神经网络模型对应测试集的校正决定系数 $R_{adjusted}^2$ 平均值为 0.70,拟合效果可以接受。特别是,模型 4 对应测试集的校正决定系数约为 0.91,均方根误差为 19.8,拟合效果最为显著。模型 4 对应测试集的拟合效果如图 1 所示(因测试集样本量较大,全面展示将使图片清晰度不佳,因此仅选取部分样本量的拟合效果予以展示)。从图 1 中可以看出,大部分样本的预测值与实际值都非常接近,说明用该模型预测论文发表 10 年后的累计被引次数是非常有效的。

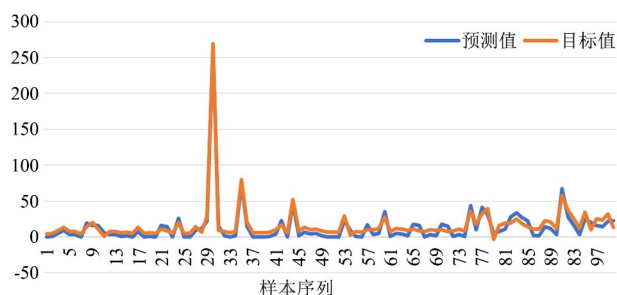


图 1 神经网络模型对应部分测试集样本量的拟合结果

Fig. 1 Fitting Results of Neural Network Model to Corresponding Portion of Test Set Samples

一方面,本研究在选择特征时综合了与论文创新力和影响力相关的多方面特征,且模型目标值选择论文发表 10 年后的累计被引次数,由此构建的模型用来评价学术论文质量或科学价值是科学的;另一方面,本研究所构建的模型特征空间中,每个特征值的计算都最晚截止于论文发表后两年,特别是反映论文短期影响力的特征只选择了论文发表后两年内获得的累计被引次数。因此可以用于对发表短期内成果的质量、贡献或影响进行较为真实的判断和衡量;最后,本研究选择 BP 神经网络算法进行模型训练,其很强的自学习和自适应能力,可以用于高效解决复杂的非线性拟合问题,相较于传统权重确定方法及其他很多机器学习算法,BP 神经网络算法的拟合效果更优。因此,本研究所构建的基于 BP 神经网络模型的论文评价模型是合理的、科学的,具有很强的应用价值。

5 结 论

学术论文的科学价值自其出版之日起就已形成,但只有随长时间的科学发展其科学价值才能得以充分体现。而在国内外科学事业的许多管理与决策中,为衡量学者当前的科研水平与学术贡献,通常特别注重对学者最近几年内发表论文科学价值的评价。为此,本研究尝试基于论文发表两年内的特征(有些特征甚至早于论文发表前)并采用不同的机器学习算法构建适用于在论文发表早期评价其科学价值的模型。研究表明,基于论文发表两年内的特征可以训练出较为精确地评价论文科学价值的模型,且训练出的神经网络模型的拟合效果非常显著。此外,尽管学术论文的科学价值是由其研究内容决定的,但本研究证明通过论文的关联特征也可以较为准确

地对其进行评价,包括论文自身特征、期刊特征、作者特征和引文特征。在评价学术论文时,这些特征并非单独起作用,而是综合且非线性地起作用。

本研究也有一定的局限性。第一,本研究将论文发表 10 年后的累计被引次数作为评价论文科学价值的目标值,无法完全等价于论文的科学价值。后续将探索采用无监督机器学习算法构建论文评价模型,以提高评价结果的说服力。第二,在构建评价模型时只针对图书情报学科论文进行了研究,因此研究结论并非一定适用于所有学科,特别是自然学科。后续将继续尝试在其他学科上做试验,力求构建出适用于所有学科或至少在多个社会科学学科中可通用的论文评价模型。

参 考 文 献

- [1] 田贤鹏. 高校教师学术代表作制评价实施: 动因、挑战与路径 [J]. 中国高教研究, 2020, (2): 85-90.
- [2] Raan A F J. Sleeping Beauties in Science [J]. Scientometrics, 2004, 59 (3): 467-472.
- [3] Seglen P O. Why the Impact Factor of Journals Should Not Be Used for Evaluating Research [J]. BMJ, 1997, 314 (7079): 498-502.
- [4] 刘明. 同行评议刍议 [J]. 科学学研究, 2003, 21 (6): 574-580.
- [5] Yogatama D, Heilman M, O'Connor B, et al. Predicting a Scientific Community's Response to an Article [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011: 59-604.
- [6] 杨建林, 钱玲飞. 基于关键词对逆文档频率的主题新颖度度量方法 [J]. 情报理论与实践, 2013, 36 (3): 99-102.
- [7] Yu T, Yu G, Li P, et al. Citation Impact Prediction for Scientific Papers Using Stepwise Regression Analysis [J]. Scientometrics, 2014, 101: 1233-1252.
- [8] Gazni A, Didegah F. Investigating Different Types of Research Collaboration and Citation Impact: A Case Study of Harvard University's Publications [J]. Scientometrics, 2011, 87: 251-265.
- [9] Liu C L, Sheng N H. A Study on the Influencing Factors of Cited Times and Social Attention Score of Data Science Papers [J]. Data Science and Informetrics, 2022, 2: 55-67.
- [10] Yan R, Tang J, Liu X, et al. Citation Count Prediction: Learning to Estimate Future Citations for Literature [J]. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 1247-1252). Presented at the CIKM '11: International Conference on Information and Knowledge Management, Glasgow Scotland, UK: ACM, 2011.
- [11] Ma A, Liu Y, Xu X, et al. A Deep-learning Based Citation

- Count Prediction Model with Paper Metadata Semantic Features [J]. *Scientometrics*, 2021, 126 (8): 6803–6823.
- [12] 任海英, 王德营, 等. 主题词组合新颖性与论文学术影响力的关系研究 [J]. *图书情报工作*, 2017, (9): 87–93.
- [13] Bosquet C, Combes P. Are Academics Who Publish More Also More Cited? Individual Determinants of Publication and Citation Records [J]. *Scientometrics*, 2013, 97: 831–857.
- [14] Weis J W, Jacobson J M. Learning on Knowledge Graph Dynamics Provides an Early Warning of Impactful Research [J]. *Nature Biotechnology*, 2021.
- [15] Abrishami A, Aliakbary S. Predicting Citation Counts Based on Deep Neural Network Learning Techniques [J]. *Journal of Informetrics*, 2019, 13: 485–499.
- [16] 赵蓉英, 郭凤娇, 谭洁. 基于 Altmetrics 的学术论文影响力评价研究——以汉语言文学学科为例 [J]. *中国图书馆学院*, 2016, 42 (1): 96–108.
- [17] Ruan X, Zhu Y, Li J, et al. Predicting the Citation Counts of Individual Papers Via a BP Neural Network [J]. *Journal of Informetrics*, 2020, 14 (3): 101039.
- [18] Wang M, Yu G, An S, et al. Discovery of Factors Influencing Citation Impact Based on a Soft Fuzzy Rough Set Model [J]. *Scientometrics*, 2012, 93: 635–644.
- [19] Rumelhart D E, Hinton G E, Williams R J. Learning Representations By Back-propagating Errors [J]. *Nature*, 1986, 323 (6088): 533–536.
- [20] Wong T C, Chan A H S. A Neural Network-based Methodology of Quantifying the Association Between the Design Variables and the Users' Performances [J]. *International Journal of Production Research*, 2015, 53 (13): 4050–4067.
- [21] Lee S, Choeh J Y. Predicting the Helpfulness of Online Reviews Using Multilayer Perceptron Neural Networks [J]. *Expert Systems with Applications*, 2014, 41 (6): 3041–3046.
- [22] 贺婉莹. 基于机器学习的论文学术创新力评价研究 [D]. 南京: 南京大学, 2019.
- [23] 胡译文, 周西姬. 基于 BP 神经网络和 MIV 算法的高价值专利预测与影响因素分析 [J/OL]. *信息资源管理学报*. <https://link.cnki.net/urlid/42.1812.g2.20231025.2230.004>.

(责任编辑: 郭沫含)

(上接第 159 页)

- [5] Fresneda J E, Gefen D. Gazing at the Stars is Not Enough, Look at the Specific Word Entropy, Too! [J]. *Information & Management*, 2020, 57 (8): 103388.
- [6] 卢国强, 黄微, 刘毅洲. 群体极化视域下突发事件网络舆情极端观点识别研究 [J]. *情报资料工作*, 2023, 44 (1): 42–51.
- [7] Yang C, Gao B, Yang J. Option Pricing Model with Sentiment [J]. *Review of Derivatives Research*, 2016, 19 (2): 147–164.
- [8] Smales L A. Asymmetric Volatility Response to News Sentiment in Gold Futures [J]. *Journal of International Financial Markets, Institutions & Money*, 2015, 34: 161–172.
- [9] Gong X, Lin B. The Incremental Information Content of Investor Fear Gauge for Volatility Forecasting in the Crude Oil Futures Market [J]. *Energy Economics*, 2018, 74: 370–386.
- [10] Lin C B, Chou R K, Wang G H K. Investor Sentiment and Price Discovery: Evidence from the Pricing Dynamics Between the Futures and Spot Markets [J]. *Journal of Banking & Finance*, 2018, 90: 17–31.
- [11] Seo S W, Kim J S. The Information Content of Option-implied Information for Volatility Forecasting with Investor Sentiment [J]. *Journal of Banking & Finance*, 2015, 50: 106–120.
- [12] Blei D M, Ng A Y, Jordan M I, et al. Latent Dirichlet Allocation [J]. *Journal of Machine Learning Research*, 2003, 3 (4/5): 993–1022.
- [13] Gao Y B, Yang L Y, Hu W J, et al. Research on Domanial Microblog Topic Evolution Based on HDP Model [J]. *Computer Engineering*, 2018, 5: 132–145.
- [14] 刘熠. 微博舆情信息受众的参与行为拟合与靶向引导研究 [D]. 长春: 吉林大学, 2020.
- [15] 庞庆华, 董显蔚, 周斌, 等. 基于情感分析与 TextRank 的负面在线评论关键词抽取 [J]. *情报科学*, 2022, 40 (5): 111–117.
- [16] 陈同辉, 鞠荣华. 中国商品期货市场价格发现能力及影响因素研究 [J]. *价格理论与实践*, 2021, (8): 118–122.
- [17] Wang S, Chen Z Y, Liu B. Mining Aspect-Specific Opinion Using a Holistic Lifelong Topic Model [C] // *International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016.
- [18] Shams M, Baraani-Dastjerdi A. Enriched LDA (ELDA): Combination of Latent Dirichlet Allocation with Word Cooccurrence Analysis for Aspect Extraction [J]. *Expert Systems with Applications*, 2017, 80: 136–146.
- [19] Mimno D M, Wallach H M, Talley E M, et al. Optimizing Semantic Coherence in Topic Models [C] // *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27–31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL*. Association for Computational Linguistics, 2011.

(责任编辑: 郭沫含)