

# 在线学习社区发帖质量评价的回归模型研究

刘金晶,王丽英

(南京师范大学教育科学学院,江苏 南京 210023)

[摘要] 在线学习社区中,多样化教学情境下基于开放话题的发帖使学生能够阐述自我知识更新的进展,但也伴随着越来越高的阅读评价解析成本。为此,可应用机器学习理论构建发帖质量评价回归模型来实现文本自动评价。首先构建文本质量评价指标及其计算所依赖的概念关系图,然后结合专家评分标准与评分结果选用多种拟合回归算法对文本质量进行预测评价,最后以拟合优度、交叉验证精度方差和均方误差为指标评估算法效果模型,测试以倡导知识建构学习理念的“数课”平台《网络安全与维护》课程的575条发帖为数据集,实现了网络安全领域的概念关系图存取和发帖质量的特征提取与预测评价。实验表明,梯度树上升回归算法的准确性、稳定性均优于其他算法模型。该回归模型能够从5个与文本质量显著相关的特征维度,即可读性、相关性、内聚度、专业度和探究度,有效地实现文本质量自动评价,从而为教师减负和学生自我诊断提供帮助。

[关键词] 机器学习,文本质量评价,概念关系图,在线学习社区,开放话题

[中图分类号] G434 [文献标志码] A [文章编号] 1672-1292(2020)01-0033-09

## Regression Model Research on Posting Quality Evaluation in Online Learning Community

Liu Jinjing, Wang Liying

(School of Education Science, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** In the online learning community, postings based on open topics in a variety of teaching contexts enable students to articulate the progress of self-knowledge update, but with the increasing costs of reading, evaluating and analyzing these posts. To alleviate it, the machine learning theory is applied to construct a regression model for posting quality evaluation. Firstly, the model constructs five dimensions of text quality evaluation and the conceptual relationship diagram on which the calculation depends. Secondly, combined with the expert scoring standard and the scoring result, the model uses multiple fitting regression algorithms to predict and evaluate the text quality. Finally, the algorithm is evaluated by coefficient of determination, cross-validation accuracy and mean square error. This paper takes 575 posts of the “Shu Ke” platform which advocates knowledge construction from “Network Security and Maintenance” course as a dataset. The model can realize the feature extraction and predict the posting quality evaluation by means of conceptual relationship diagram. The experimental results show that the accuracy and stability of the gradient decision boosting tree regression algorithm are better than those of other algorithm models. The regression model can effectively realize the automatic evaluation of text quality based on five characteristic dimensions that are significantly related to text quality: readability, relevance, cohesion, professionalism and exploration, thus reducing the burden on teachers and helping students with self-diagnosis.

**Key words:** machine learning, text quality evaluation, conceptual relationship diagram, online learning community, open topic

《中国教育现代化2035》的印发启示着更加智能的教育现代化对教育质量发展的战略作用。以尊重学生认知发展规律为第一要素、提高学生深度学习水平的教学理念和教学方法不断被提出,学生的学习效果逐渐以学生的消化吸收思考的结果——能够反映批判理解、信息整合、知识建构水平的文本<sup>[1]</sup>来呈现。在线学习社区中的文本不再仅限于有标准答案的选择题、简答题和参考答案的论述题,还逐渐囊括了开放话

收稿日期:2019-05-26.

基金项目:全国教育科学“十三五”规划教育部重点课题(DIA170375).

通讯作者:王丽英,博士,讲师,研究方向:计算机图形学、信息技术. E-mail: wangliying@njnu.edu.cn

题的知识总结与自我探究,如毕博平台讨论版、MOOC平台讨论区、数课云平台等等。开放话题的发帖作为学习者学习过程的数据源之一,能够发挥评价学习者学习进展与技能水平的作用;然而因其具有主观性、发散性、无标准答案的特征,这些发帖的阅读与质量评估往往通过人工评价的方式来达成。当在线发帖数量过多时,教师便面临着较高的阅读解析成本,其原因在于:一是文本内容过多,信息内容产生冗余、过载;二是海量文本中夹杂一定规模的复制、抄袭内容,增加了不必要的评阅负担,消耗了教师过多的时间与精力。

构建基于机器学习算法的在线学习平台的发帖质量评价模型,一方面能够减轻教师负担,较为准确地识别学生的学习情况;另一方面能够进行更加精准、更有针对性的教学,提升教学品质。这是教育智能化的一次大胆尝试,也是构建教育信息化2.0时代的新生态的重中之重<sup>[2]</sup>。

## 1 相关研究现状

机器学习算法在教育领域中的文本内容分析方面有着广泛的应用,例如论坛话题评价、教学资源挖掘、主观试题批改等。其中,分类或回归模型属于有监督的机器学习算法,分类模型适用于离散类别预测,而回归模型适用于根据若干个解释变量对连续类型的因变量进行预测。常用的有监督学习模型主要有线性、决策树、贝叶斯、支持向量机、神经网络等。常用的回归模型包括线性支持向量机回归、多项式回归、岭式回归、一般线性回归、多层感知器回归、梯度树上升回归、高斯过程回归等。

### 1.1 不同类型文本的评价研究

在教育论坛上,文本内容从话题上可分为课程内容相关、课程设置安排相关及与课程完全无关的讨论<sup>[3]</sup>,目前针对文本内容评价学习效果的研究还较少。侯明良<sup>[3]</sup>根据主题义原之间的7种语义关系计算文本相似性,在此基础上对MOOC讨论区中发帖的主题聚类;同时半自动构建知识语义网实现导航;通过学习者主观检验,基于学习兴趣相似度推荐学习伙伴的准确率达到70.4%。Au等<sup>[4]</sup>提出了开放式的问题来获得学生在使用MOOC课程后的反馈,并使用文本挖掘技术生成基于反馈的概念图。史文祥<sup>[5]</sup>提出基于LDA的文档主题行为模型,研究StarC讨论区中学习者交互行为特征、讨论主题与学习成效之间的关联关系,采用逻辑回归和支持向量机预测学习成绩,精度达到73%。李敏<sup>[6]</sup>以知识建构层次模型对Coursera讨论区的发帖进行内容编码与归类整理,以此分析社区成员的知识建构水平。张平霞<sup>[7]</sup>以布鲁姆的认知分类法为理论基础对MOOC讨论区的发帖进行人工分类,再运用文本向量和朴素贝叶斯分类器自动评价文本的认知水平,在测试集上的分类预测准确率达到85.3%。

国外学者对主观题自动评分系统的研究起步较早。Valenti等<sup>[8]</sup>对国外相对成熟的8项评分系统,即项目作文评分(PEG)、智能作文评价(IEA)、电子作文评级(E-Rater)、概念评级(C-Rater)、教育测试服务(ETS I)、贝叶斯作文测试评分系统(BETSY)、智能作文评分系统(IEMS)、自动评分(Automark),进行了比较分析,并从技术及精确度、多元回归相关性及系统评分结果与人工评分结果之间的符合程度百分比3方面比较这些系统的性能。国内学者对主观题评分系统进行了改进。谭冬晨<sup>[9]</sup>采用词性语义相似度及若干浅层文本相似度作为文本的特征,构建决策树,完成主观题自动评分。王漪<sup>[10]</sup>模拟教师评阅主观题的思维,引入模糊数学理论中的单向贴适度模型,实现了简答题的自动评分;依赖知网构造文本相似度矩阵模型和由此形成的最大值序列逐层实现文本语义相似度计算,从而实现论述题的自动评分。

在论文自动评估领域,Shehab等<sup>[11]</sup>提出了一种由写作特征分析工具和神经网络评分引擎两个互补部分组成的自动论文评分系统(AEGS)的混合方法,从而实现学生论文的自动评分。Ramalingam等<sup>[12]</sup>将文本实体的语料库分类为少量的离散类别对应可能的分数,利用机器学习技术开发了一种自动化的论文评估系统。Liu等<sup>[13]</sup>对以化学、电气和计算机科学工程专业为代表的中国工科学生的英文论文进行自动评价,提出了一种新的评分算法——PSO-SVR,并证明该算法优于传统的多元线性回归、回归支持向量机和K近邻算法,且在预测不规则数据集的评分方面具有较强的鲁棒性。

### 1.2 文本质量评估的特征选取

文本质量评估的核心是寻找良好的数据质量特征维度构建方法,即特征选取。钟将等<sup>[14]</sup>从相关度、全面度、专业度、内聚度及可读性5个质量维度来衡量评论文本的质量,构建了句法分析、LDA主题分析、FCA分析到概念格约简的算法框架,实验表明该模型的预测精度显著高于基于文本分析方法的预测模

型. 靳健等<sup>[15]</sup>用语言结构特征、产品特征、基于信息质量的特征和基于信息论的特征来描述评论的质量,并提出 Co-training 算法分析评论质量,通过与 SVM 和朴素贝叶斯算法进行对比来验证算法的可靠性. 王洪伟等<sup>[16]</sup>以信息采纳模型为理论依据,采用语言学特征、语义特征、信息特征、评论者相关特征来衡量评论质量,并采用梯度提升的决策树 GBDT 模型自动识别高质量评论的关键影响因素. 聂卉<sup>[17]</sup>提出信息含量、主题内容关联及情感表达是评论内容被感知具有效用价值的重要因素,词量、主题相关度和情感类型是衡量文本特征属性相对重要的因素;其构建基于 TFIDF 权值的文本特征矢量,采用余弦相似度计算评论内容与权威的官方评测产品描述的语义相似度,获取面向语义内容的主题相关性特征指标. 张艳丰等<sup>[18]</sup>从评论属性的形式特征和内容特征两方面构建影响评论效用的 6 项属性指标,对在线评论语句进行因素划分和句法分析,根据“在线评论效用”取值及评论长度、产品属性词、情感强度、修饰词数量 4 个指标的量化值,构建模糊神经网络系统. 王忠群等<sup>[19]</sup>建立商品评论深刻性模型,评价指标包括:常见功能特征词量、关键功能特征词量、特征词汇量、功能特征词分布层次等,并通过专家构建概念树、算法实现评论深刻性排序、问卷调查验证来进行研究. Aika 等<sup>[20]</sup>从概念的角度进行验证,发现商品评论对消费者的帮助度与评论的每句话包括的概念数量、每条评论包括的平均概念数量相关. Nikolaos 等<sup>[21]</sup>通过对英国亚马逊商品平台上 3 万条数据进行测试,发现评论文本的可读性和评论的等级直接影响评论的帮助度,评论文本的长度直接受到评论等级分值的影响.

综上所述,现有文本质量评价的指标已考虑了修辞、句式、情感等方面的指标,但不能直接用于发帖质量的评价. 同时,回归模型在文本质量评价预测的准确度还不高,有进一步提高的可能. 本文构建在线学习社区开放话题发帖质量人工评价标准和机器学习评价指标体系,运用概念图及回归算法,构建在线学习社区发帖质量的预测评价模型,并验证其准确性和稳定性.

2 在线学习社区发帖质量评价回归模型的研究框架

根据有监督的机器学习方法流程,在线学习社区发帖质量评价模型的研究框架如图 1 所示,主要包括 5 个部分:数据采集及处理、特征提取、模型构建、模型验证、效果评估.

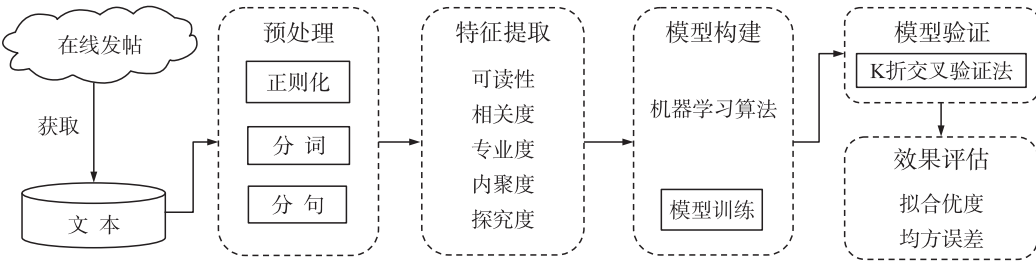


图 1 在线学习社区发帖质量评价研究框架

Fig. 1 Research framework of posting quality evaluation in online learning community

2.1 数据集描述及处理

以“数课”平台《网络安全与维护》课程中的学生发帖作为数据集,数据采集时间为 2017 年 11 月至 2018 年 1 月,获取其发帖时间、标题和内容,数据集累计 633 条. 由于发帖是其作业必不可少的组成部分,因此不进行去冗处理,以最大程度保留学生作业的真实性与完整性. 经过数据去重、去除 HTML 标记等预处理,剩余具有可读性的文本数据集总计 575 条.

2.2 人工评分标准设计与协同分析

通过对 3 位专家进行访谈,设计 5 个维度对数据集进行评分考量,评分包括 6 个等级,即 0~5,如表 1 所示. 人工评分标准的各个维度含义如下:

可读性:文通句顺,自然流畅,段落分明,格式清楚.

相关性:标题及发帖内容的主题与网络安全领域的内容紧密相关.

表 1 人工评分简单标准

Table 1 Simple standard of manual scoring

评分	标准
0	无相关性
1	可读性差,相关性差
2	可读性较差,相关性较差,逻辑性较差
3	可读性一般,相关性一般,逻辑性一般,内容完整
4	可读性较强,相关性较强,逻辑性较强,有一定深度,探究性较强
5	可读性强,相关性强,逻辑性强,很有深度,探究性强



逻辑性:思路清晰,前后内容关联性强,条理分明,论述严谨,科学合理.

深度:专业性强,深邃、有思想,剖析问题鞭辟入里,能够揭示问题的本质.

探究性:体现出对问题解决的分析、探究、总结过程以及结论、方案的发现等.

3 位专家按照已制定的人工评分标准进行评分,取每条帖子的平均分作为数据集的评分结果,以保证评分的全面性、一致性. 使用 SPSS 工具中的 Kendall 相关系数检验 3 位专家的评分是否协同一致, Kendall 相关系数  $r$  分别为 0.600、0.666、0.788,  $p < 0.001$ , 表明相关性显著,其可信度是中等偏高的.

2.3 特征提取

学生发帖限定于网络安全领域,因此评价发帖质量时需要以网络安全领域的典型特征为依据. 文献 [19] 依据商品广告、商品说明书和编辑商品测评等构建商品特征概念树,以此为基础度量评论深刻性. 借鉴该方法,构建网络安全领域概念关系图用于检索发帖内容中提及的专业术语,从而进行特征提取和计算,自动提取特征的实施路径如图 2 所示.

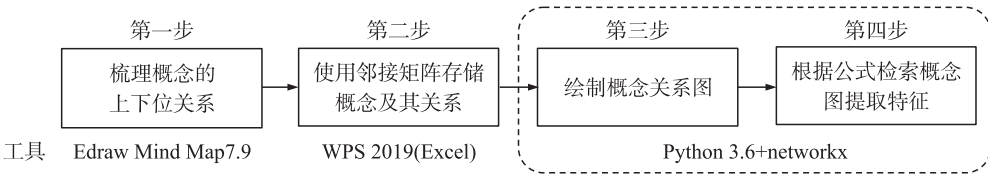


图 2 自动提取特征实施路径

Fig. 2 Implementation paths on automatic extraction of feature

2.3.1 概念关系图的构建与存储

以黄传河等编写的《网络安全》<sup>[22]</sup>一书为基础梳理网络安全领域的专业术语,再辅以专家建议、其他参考书目等进行补充,构建网络安全领域的概念关系图,如图 3 所示. 其存储结构采用邻接矩阵表示. 图中节点表示概念;有向边表示概念的上下位关系;根节点为总概念,称为第 0 层的概念;图的 1~ $n$  层为下位概念.

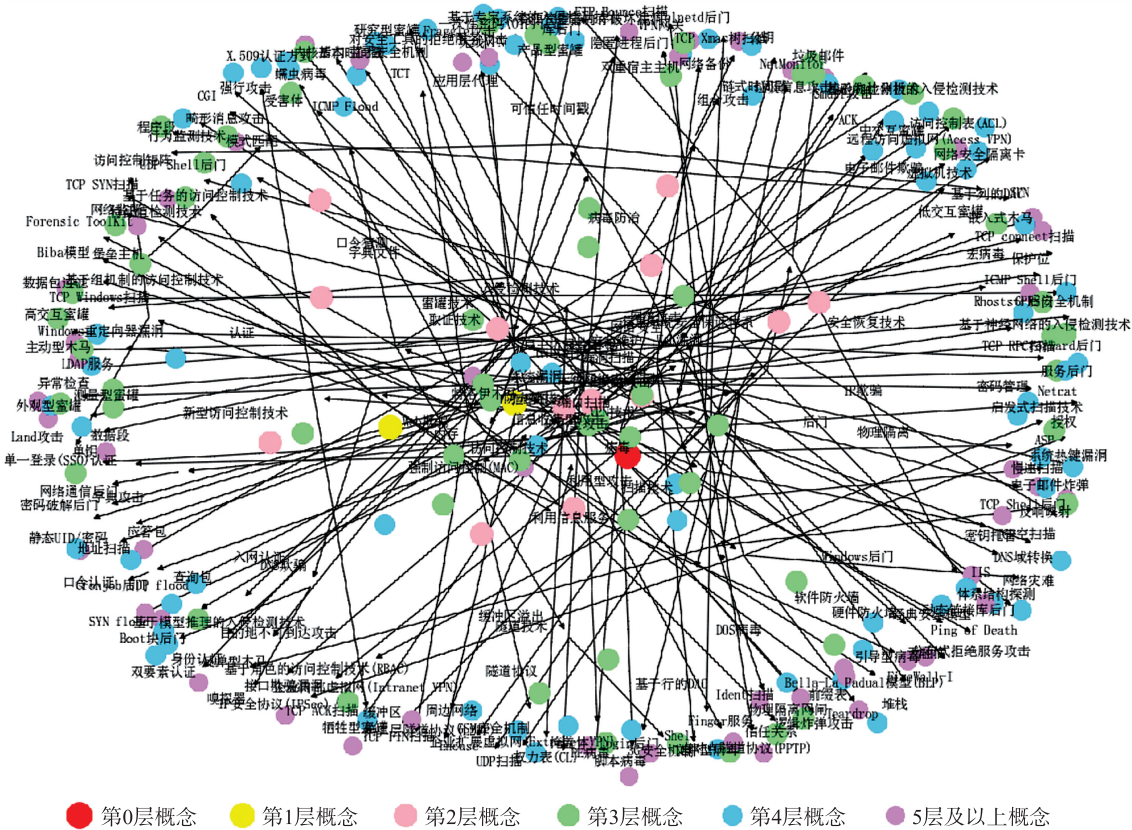


图 3 使用 networkx 绘制的概念关系图

Fig. 3 Conceptual relationship diagram with networkx

### 2.3.2 特征指标的假设、计算与验证

基于数据集特定的教学情境以及文献[14]的设计,假设在线学习社区发帖质量与表2所示的5个维度的指标相关,其中每一个指标的具体特征项将根据概念关系图进行量化评分。

表2 在线发帖质量评价指标阐释

Table 2 Explanation of evaluation indicators of online posting quality

假设	指标	定义	具体特征项
1	可读性	用语简洁、规范、无冗余	有效字符数
2	相关度	发帖标题及内容与网络安全领域的相关程度	特征词频次
3	内聚度	发帖内容之间前后的关联度	特征词分布集中度
4	专业度	阐述问题/现象/原理、揭示问题本质的深度	特征词分布路径的平均长度
5	探究度	对问题分析、探究、总结的程度	具有探究性词语的词量

假设 TC(Text Content)为发帖文本的内容,TP(Text Phrase)为发帖文本分词后的短语对象集合,由于标点符号也是影响文本质量的因素,因此 TP 集合中不去除标点符号, $n=|TP|$ ,CS(Concept Set)为关联的网络安全领域的概念集合。各指标分别按公式计算,所得分值根据表3映射成相应级别。

表3 评价指标分值对应表

Table 3 Feature values of evaluation indicators and scores

分值	指标				
	可读性	相关度	内聚度	专业度	探究度
0	0	0%	0	0	0%
1	1	0%~5%	0~0.2	1	0%~1%
2	2	5%~10%	0.2~0.4	2	1%~2%
3	3	10%~20%	0.4~0.6	3	2%~3%
4	4	20%~30%	0.6~0.8	4	3%~4%
5	5	>31%	>0.8	5	>4%

(1)可读性(readability):可读性较高的文本通常字符数量较多。由于部分发帖内容中仅重复发帖标题,而标题通常不超过20个字符,因此将字符数作为度量可读性的特征项,公式为:

$$\text{read}(\text{TC}) = \begin{cases} 0, & \text{length}(\text{TC}) \leq l_0; \\ k, & \text{length}(\text{TC}) \in (l_{k-1}, l_k]; \\ k+1, & \text{length}(\text{TC}) > l_4. \end{cases} \quad (1)$$

式中, $l_k$ 可根据实际调整,本文采用的 $l_k$ 分别取值为20,50,100,150,200;其中 $k=0, \dots, 4$ 。函数length为获取到的文本字符数,包括文本中出现的标点符号。

(2)相关度(relativity):发帖的主题相关性越高,则文本中提及的专业术语越多。因此相关度根据发帖中具有网络安全领域关联特征的词语在所有词语中所占的比例来度量,公式为:

$$\text{rela}(\text{TP}) = \frac{\sum_{tp_i \in \text{TP}} f(tp_i)}{n}, \quad (2)$$

式中, $tp_i$ 为TP中的短语; $n$ 为所有文本短语的数量之和。函数 $f$ 的定义为:

$$f(tp_i) = \begin{cases} 1, & tp_i \in \text{CS}; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

例如,某学生发帖中出现了“物理隔离”一词,而该词属于网络安全领域的概念,则 $f=1$ ;而对于学生发帖中出现的“日期”一词, $f$ 赋值为0;以此类推,对某一发帖中的所有文本短语与概念关系图的节点进行匹配,匹配成功赋值为1,匹配失败赋值为0,函数rela的值即为 $f$ 的累加总和与文本短语数量的比值。

(3)内聚度(cohesion):深度不同的发帖提及的主要专业术语在概念图中具有不同的位置,通过这些专业术语的平均位置来衡量发帖的总体内聚程度。因此,内聚度根据发帖中出现的概念与概念关系图的根节点的最短路径长度之和的平均数来度量,公式为:

$$\text{cohe}(\text{TP}) = \frac{\sum_{i=1}^n \text{depth}(tp_i)}{n}, tp_i \in \text{CS}, \quad (4)$$

式中,函数  $\text{depth}$  的值为某一概念与概念图的根节点之间的最短距离; $n$  为发帖文本中所有词语数量之和.例如,某学生发帖中出现了“拒绝服务”一词,而网络安全领域概念关系图中的根节点为“网络安全”,根据式(4)进行测算,两者之间的最短路径长度为 4,则该文本此概念的  $\text{depth}$  值为 4;同理,对某一发帖中出现的所有概念测算其与根节点之间的最短路径距离,距离之和与文本短语数量的比值即为此时函数  $\text{cohe}$  的值.

(4) 专业度( $\text{specificity}$ ):发帖中提及的专业术语在概念层次上所占的层次越高,则专业性越强.因此,专业度根据发帖涉及网络安全领域术语的最大深度来衡量,公式为:

$$\text{spec}(\text{TP}) = \max(\text{depth}(tp_i)), tp_i \in \text{CS}. \quad (5)$$

例如,一段发帖中出现了“拒绝服务”、“物理隔离”、“身份认证”这 3 个概念,根据式(5)测算,这 3 个概念的  $\text{depth}$  值分别为 4、3、4,则此时  $\text{spec}$  函数的值为其中的最大值 4.

(5) 探究度( $\text{exploration}$ ):探究性较强的发帖常常伴有明显的观点表达特征,因此探究度根据以第一人称作为主语的探究性、总结性的词句占比来度量,特征词包括“我”、“发现”、“想”、“探讨”、“感受”等,由于人们也常常将自己想要研究的问题以问句的形式来呈现,因此标点符号“?”也包含在特征符号集合之内,所有特征词集合记为  $S(\text{symbol})$ . 集合  $S = \{\text{“我”}, \text{“发现”}, \text{“想”}, \text{“探讨”}, \text{“感觉”}, \text{“觉得”}, \text{“认为”}, \text{“思路”}, \text{“总结”}, \text{“?”}\}$ ,计算公式为:

$$\text{expl}(\text{TP}) = \frac{\sum_{tp_i \in S} g(tp_i)}{n}, \quad (6)$$

式中, $n$  为发帖文本中所有词语数量之和;函数  $g$  的定义为:

$$g(tp_i) = \begin{cases} 1, & tp_i \in S; \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

例如,某一发帖内容为“我认为系统漏洞在一定程度上是无法避免的,只能通过发现漏洞——规避漏洞的方法来解决.接下来我将对漏洞的形成原理做一个总结.”这段话中符合集合  $S$  的特征词有“我”“认为”“发现”“总结”,累计出现频次为 5;该文本经分词处理后的内容为 $\{\text{“我”}, \text{“认为”}, \text{“系统漏洞”}, \text{“在”}, \text{“一定”}, \text{“程度”}, \text{“上”}, \text{“是”}, \text{“无法”}, \text{“避免”}, \text{“的”}, \text{“,”}, \text{“只能”}, \text{“通过”}, \text{“发现”}, \text{“漏洞”}, \text{“—”}, \text{“一”}, \text{“规避”}, \text{“漏洞”}, \text{“的”}, \text{“方法”}, \text{“来”}, \text{“解决”}, \text{“.”}, \text{“接下来”}, \text{“我”}, \text{“将”}, \text{“对”}, \text{“漏洞”}, \text{“的”}, \text{“形成”}, \text{“原理”}, \text{“做”}, \text{“一个”}, \text{“总结”}, \text{“.”}\}$ ,词语数量为 37,此时函数  $\text{expl}$  的值为  $5/37 \approx 0.13 > 0.04$ .

部分数据集的特征指标分值如表 4 所示.在此数据集上,学生发帖呈现出可读性最高、专业度次之、内聚度最低的特点.结合学生发帖的主题与概念关系图的比较,可见学生对于研究问题的横向扩展和纵向延伸还不够深入.发帖的指标分值分布表现出较为明显的学生个性特征,文本篇幅、行文思路与语言组织、反思性话语表达等均在一定程度上反映出学生在平时作业习惯、说话风格、逻辑组织上的差异.

表 4 发帖内容的各项指标分值与人工评分均值(部分)

Table 4 Partial indicators scores and manual scores of posting content

发帖编号	公式提取的指标分值					人工评分均值
	可读性	相关度	内聚度	专业度	探究度	
1	4	0	0	0	0	1.67
2	5	1	1	4	0	3.00
3	4	0	0	0	0	2.33
4	2	0	0	0	5	1.67
5	5	0	0	0	0	3.00
6	3	0	0	0	0	2.33
7	2	0	0	0	0	2.33
8	5	1	1	4	0	2.33
9	1	0	0	0	0	0.33
10	5	1	1	3	2	2.67
...	...	...	...	...	...	...

对所有指标分值与人工评分均值使用 Spearman 相关系数进行相关性分析.结果表明,可读性、相关度、内聚度、专业度、探究度与人工评分均值的相关系数分别为 0.901、0.427、0.439、0.482、0.317,相关系数的显著性均为 0.000,小于 0.01,说明以上 5 个指标与发帖质量的相关性是高度显著的.由此可以解释与发



帖质量相关的 5 个假设均成立,因而进行机器学习的 5 个指标均为有效特征.

2.4 发帖质量评价的回归模型构建

根据上文特征提取的指标分值,进行机器学习训练的基本思路如图 4 所示. 已知  $X = [x_{ij}]_{mn}$ ,  $Y = [y_i]_m$ , 其中  $m$  为数据集的数量 575,  $n$  为指标数量 5, 即每个发帖有 5 个特征指标分值, 构成  $x_i$ . 人工评价分值的平均分作为  $y_i$ , 得分在 0~5 之间.

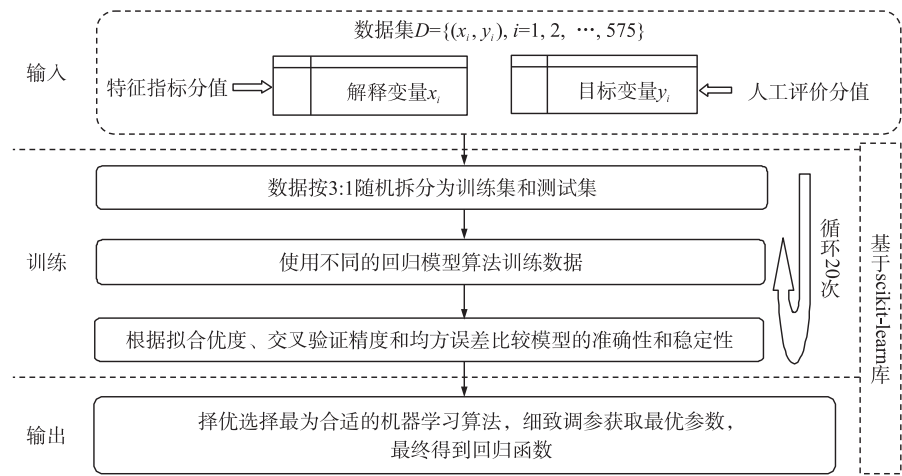


图 4 机器学习算法选择路径

Fig. 4 Path of selecting optimal machine learning algorithm

2.5 模型验证及效果评估

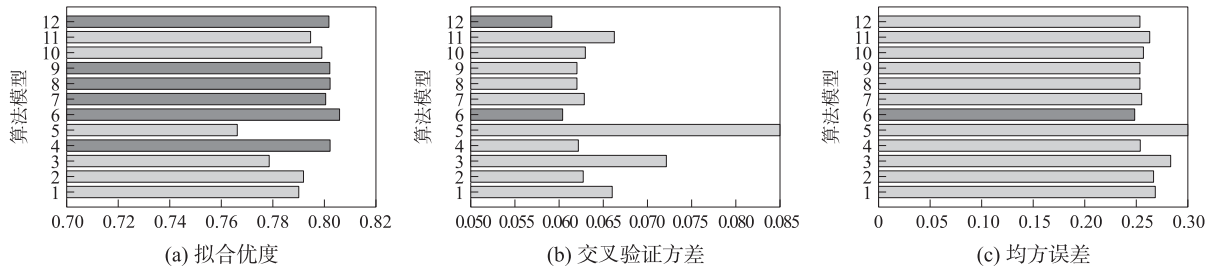
采用 20 次 5 折交叉验证法来验证模型稳定性,也即对模型的验证集性能取方差,用于判定不同学习算法平均性能之间的微小差异是否有意义. 通过在测试集上计算拟合优度  $R^2$  和均方误差来评估模型的准确性.  $R^2$  是数据集的回归平方和 (sum of squares for regression, SSR) 与总离差平方和 (total sum of squares, SST) 的比值, 其中, SSR 是指因变量的回归值与其均值的差的平方和, SST 是指每个因变量的实际值与其均值的差的平方和,  $R^2$  的取值范围为  $[0, 1]$ , 越接近 1 则拟合程度越好. 均方误差是每条数据记录的模型预测值与目标变量真实值之差的平方和的均值, 均方误差越小说明数据的变化程度越小. 根据 20 次训练的拟合优度、交叉验证方差和均方误差的平均值择优选择模型.

3 发帖质量评价的回归模型效果分析

按照 3:1 切分训练集和测试集能使模型训练的效果最佳,切分后的训练集数量为 431, 测试集数量为 144.

3.1 回归模型初步筛选

使用 scikit-learn 库中所有回归算法的回归器对训练集进行训练,在默认参数下算法的准确性和稳定性评估如图 5 所示. 从图 5(a) 可以看出,拟合优度大于 0.8 的算法有高斯过程回归、梯度树上升回归、多层感知器回归、一般线性回归、岭式回归、多项式回归、线性支持向量机回归,这些算法在测试集上的预测



1. 支持向量机回归 2. Nu支持向量机回归 3. 内核岭回归 4. 高斯过程回归 5. 决策树回归 6. 梯度树上升回归 7. 多层感知器回归 8. 一般线性回归 9. 岭式回归 10. 随机梯度下降回归 11. 多项式回归 12. 线性支持向量机回归

图 5 不同回归算法评估效果比较

Fig. 5 Comparison for evaluation effects of different regression algorithms

结果与专家评分的实际值比较接近.从图5(b)可以看出,模型中交叉验证精度方差较小的有梯度树上升回归、线性支持向量机回归,方差值在0.06附近,这两种算法的预测结果较为稳定,较少受到数据集变化的影响.从图5(c)可以看出,梯度树上升回归算法的均方误差小于0.25,在测试集上进行预测时专家评分实际值与预测值之间的平均误差较小,准确性更高.

综上所述,经过模型初步筛选环节,梯度树上升回归算法的拟合效果最好,准确性最高,且模型的性能最为稳定.

### 3.2 梯度树上升回归算法

梯度树上升回归算法主要结合回归树和提升树的思想,利用残差梯度来优化回归树的集成过程,是一种迭代的决策树算法.在该算法模型中影响最大的两个超参数为学习率和回归树的棵树,即迭代次数.已有研究表明,当回归树的棵数固定时,学习率增大往往预测精度提升更快,但较低的学习率却能达到更好的预测精度,也即学习的效果更好<sup>[23]</sup>.

逐一调整参数获取最优超参数值如下:迭代次数  $n\_estimators$  为100,学习率  $learning\_rate$  为0.1,最大叶子深度  $max\_depth$  为1,损失函数  $loss$  赋值为‘ls’,此时计算得出的模型拟合优度为0.813 58.真实值与预测值的比较如图6所示,预测值集中分布在真实值的大部分分布范围内,只有少数较高的真实值在预测范围之外.因此,基于梯度树上升回归算法模型预测的发帖评分较为准确,但对于实际得分较高的数据预测存在一定的误差,这可能是由于受到人工评分时主观因素的影响.

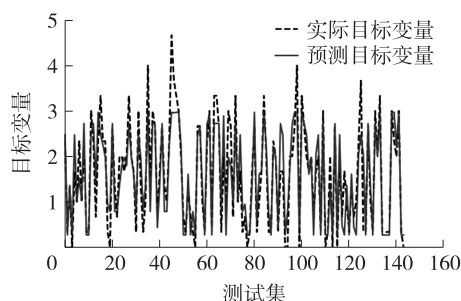


图6 发帖评分的预测值与真实值

Fig.6 Predicted and true scores of posting evaluation

## 4 结论

本文提出的回归模型适用于特定教学情境下的基于在线学习社区的开放话题发帖质量自动评价.开放话题发帖质量的人工评价简单标准与发帖质量评价特征指标体系及概念关系图依赖的计算方法具有一致性、相关性;可读性、相关度、内聚度、专业度、探究度是衡量开放话题发帖质量的5个有效特征.根据回归算法的拟合优度、交叉验证精度方差和均方误差这3个指标能够优选出准确性和稳定性均最高的梯度树上升回归算法作为在线学习社区发帖质量评价的模型.限于实验所用的数据量不够庞大,发帖质量的评价指标体系采用离散的6个等级,导致评分结果的高低优劣分界并不明显,且专家在评分过程中主观性较强,导致部分数据的预测存在误差.未来可以持续收集数据,探索学生发帖的个性化特征,提供知识建构实时导航.

### [参考文献](References)

- [1] 何玲,黎加厚.促进学生深度学习[J].现代教学,2005(5):29-30.
- [2] 杨宗凯.数据驱动个性化学习[N].中国教育报,2018-11-01(007).
- [3] 侯明良.MOOC讨论区数据挖掘与应用[D].济南:山东大学,2016.
- [4] AU C H, LAM K C, FUNG W S, et al. Using animation to develop a MOOC on information security[C]//IEEE International Conference on Industrial Engineering and Engineering Management. Bail: IEEE, 2016: 365-369.
- [5] 史文祥.基于DT-BM的学习者主题行为模型研究[D].武汉:华中师范大学,2017.
- [6] 李敏.虚拟学习社区成员互动的知识建构效果分析[D].扬州:扬州大学,2015.
- [7] 张平霞.基于文本挖掘的MOOC讨论区学习评价研究[D].重庆:重庆师范大学,2018.
- [8] VALENTI S, NEFF E, CUCCHIARELLI A. An overview of current research on automated essay grading[J]. Journal of Information Technology Education, 2003, 2: 319-330.
- [9] 谭冬晨.主观题评分算法模型研究[D].成都:电子科技大学,2011.
- [10] 王漪.文本挖掘技术的研究及其在教学平台中的应用[D].北京:北京交通大学,2014.



- [11] SHEHAB A, ELHOSENY M, HASSANIEN A E. A hybrid scheme for automated essay grading based on LVQ and NLP techniques[C]//2017 13th International Computer Engineering Conference(ICENCO). Cairo:IEEE,2016:65-70.
- [12] RAMALINGAM V V, PANDIAN A, CHETRY P, et al. Automated essay grading using machine learning algorithm[C]//2018 10th National Conference on Mathematical Techniques and its Applications(NCMTA). Kattankulathur,2018.
- [13] LIU M, WANG Y Q, XU W W, et al. Automated scoring of Chinese engineering students' english essays[J]. International Journal of Distance Education Technoloies(IJDET),2017,15(1):52-68.
- [14] 钟将,张淑芳,郭卫丽,等. 主题特征格分析:一种用户生成文本质量评估方法[J]. 电子学报,2018,46(9):2201-2206.
- [15] 靳健,季平. 用于在线产品评论质量分析的 Co-training 算法[J]. 上海大学学报(自然科学版),2014,20(3):289-295.
- [16] 王洪伟,孟园. 在线评论质量有用特征识别:基于 GBDT 特征贡献度方法[J]. 中文信息学报,2017,31(3):109-117.
- [17] 聂卉. 基于内容分析的用户评论质量的评价与预测[J]. 图书情报工作,2014,58(13):83-89.
- [18] 张艳丰,李贺,彭丽徽,等. 基于模糊神经网络的在线评论效用分类过滤模型研究[J]. 情报科学,2017,35(5):94-99,131.
- [19] 王忠群,皇苏斌,修宇,等. 基于领域专家和商品特征概念树的在线商品评论深刻性度量[J]. 现代图书情报技术,2015(9):17-25.
- [20] AIKA Q, KARIM B S S, RAM G R, et al. A concept-level approach to the analysis of online review helpfulness[J]. Computers in Human Behavior,2016,58:75-81.
- [21] NIKOLAOS K, ELENA G B, SALVADOR S A. Evaluating content quality and helpfulness of online product reviews:the interplay of review helpfulness vs. review content[J]. Electronic Commerce Research and Applications,2012,11(3):205-217.
- [22] 黄传河,杜瑞颖,张沪寅,等. 网络安全[M]. 武汉:武汉大学出版社,2004.
- [23] 龚越,罗小芹,王殿海,等. 基于梯度提升回归树的城市道路行程时间预测[J]. 浙江大学学报(工学版),2018,52(3):453-460.

[责任编辑:严海琳]