

人工智能在线教学评价方法探索

张国凯 江浩玉 卢金典

(上海理工大学光电信息与计算机工程学院, 上海 200093)

摘要: 构建科学的在线教学评价机制, 有利于实现精准的在线教学评价。近年来, 人工智能技术的发展为在线教学评价提供了新的思路, 为此本文提出一种基于人工智能技术的在线教学评价模型, 通过对在线学习平台资源教学数据进行分析, 构建基于 Bert 模型的教学智能评价模型, 预测学生的学习效果, 实现智能化教学评价。

关键词: 教学评价; 人工智能; 在线教学

在线学习打破了传统教学的时间和空间的限制, 已逐渐成为一种更加安全可靠的新型过程教学模式。与此同时, 由于在线学习与传统学习模式的差异, 依据传统教学评价方法进行在线教学评价存在一定的局限性, 特别是对于学生学习过程的行为分析, 由于不能客观地对学生学习环节进行评估, 导致在线教学的评价存在一定局限性。近年来, 人工智能技术以其高性能的表现已经在教育领域得到广泛应用。为此, 本文以人工智能技术为基础, 开展基于人工智能技术的新型在线教学评价模型与方法研究, 所构建模型旨在实现在线教学智能化效果评估, 辅助教师进行教学方案匹配以及调整。

一、研究背景

在线学习有着时间空间上的优势, 但是也存在明显的缺陷, 如信息反馈不及时、课堂互动性低等方面影响教学评价过程。具体地说, 传统的教学评价体系依赖于线下课堂的参与度数据, 而线上的学生课堂状态并不易于观测, 系统难以捕捉学生平时学习的真实情况, 这导致以成果为导向中心的传统评价体系在线上教学领域的效果有所下降, 进而无法反应课程开展的实际真实情况; 依据学生期末成绩是评价教学质量的重要部分, 但是线上课程导致学生对该课程的认知度不高, 参与度低, 互动性弱, 给出的主观评价可能带有情绪化内容, 准确性不足; 线上教学在教学过程中获得的反馈不足, 学生课堂互动状态、课后习题完成情况不利于实时统计反馈, 对于教师及时改进教学方式不利。近年来, 高速发展的科技带来了多样化的线上教学平台, 但是在给予教师网络教学的机会的同时, 不得不关注到传统的教学评价方式在匹配线上教学上产生了很多问题。因此, 如何借助人工智能技术建立一种新型智能化在线教学评价模型与方法, 实现在线教学效果评估, 缓解教师教学反馈的压力, 协助教师发现学生的问题, 并做出及时的纠正和引导, 是实现高效率线上教学评价的重要研究内容。

二、研究依据

近年来, 人工智能技术的发展促进了教学方法的转变, 特别

是其独特的算法应用, 该技术实现的在线教育系统已成为国内外社会广泛关注的焦点。基于此本文拟通过人工智能技术, 从多个不同维度及时捕获并聚焦分析学生的个人学习行为数据, 构建基于学生个人学习行为的智能化多维教学评价模型, 实现学生学习成果个性化评估以及预测。通过评估结果及时发现学生的学习问题, 并针对有风险的学生进行有针对性的干预, 提高学习成果, 教育水平以及教学效率, 改善学习技术, 学习模式, 促进在线教育的平衡发展。

三、研究方式和思路

通过收集前期多种在线学习平台资源教学数据, 如教学视频、音频、课件下载及浏览数据信息, 结合不同开放在线资源的历史学习行为数据, 对该数据进行清洗、预处理、整理和存储, 借助 Bert 预训练模型, 对数据层中各种类型数据信息进行综合分析, 获得不同教学成果画像, 将当前学生的学习状态作为输入, 通过模型输出对学生的评价, 为后续课程计划以及试练习考试分析提供数据支持, 其具体模型结构图如图 1 所示。

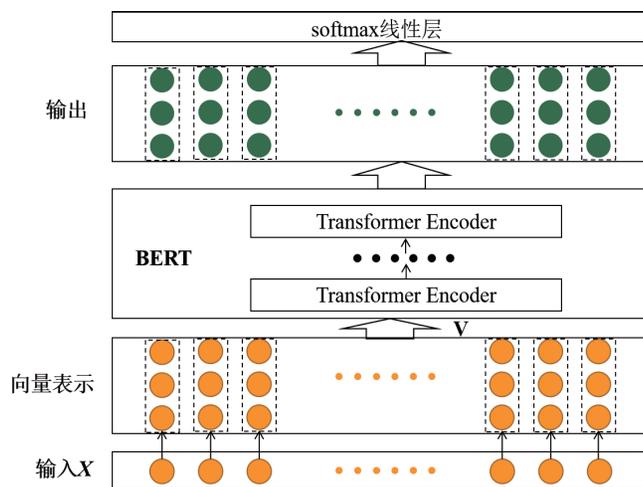


图 1 基于人工智能技术的在线教学评价模型

(一) 原始数据获取和数据预处理

为了检测学生的学习状态和课程教学质量, 需收集线上教学平台获取有关教学数据。这些数据主要由文本组成, 由于所统计的文本存在一定数量的“脏数据”, 在数据预处理之前需要对原始文本数据进行清洗。具体地, 将格式随意、缺乏语法规则的文本数据标签归一化成模型需要的格式, 并根据长度对文本进行分割和整理。本文统计的数据如表 1 所示, 依次为互动次数、课后作业完成次数、讨论区发言次数、课后作业成绩、线上课时时长, 按照各类数据的具体数值大小划分为三种评价: 优秀、一般、较差。

此外, 本文将期末成绩映射为模型的预测标签, 为了确定模型对应的输出标签将成绩按照数值大小划分为对应的标签, 对于 90-

100 分的成绩划定标签为优秀, 70-89 分的成绩划定为良好, 小于 70 分的为合格。

表 1 教学平台数据

	互动次数	课后作业完成次数	讨论区发言次数	课后作业成绩	线上课时长
1	优秀	优秀	优秀	积极	优秀
2	一般	一般	一般	一般	一般
3	较差	较差	较差	较差	较差
备注	互动次数、课后作业完成次数、讨论区发言次数: <3 次为较差, 3-5 次为一般, >5 次为优秀; 课后作业成绩 >80 为优秀, 60-80 为一般, <60 为较差; 线上课时长: 十小时以上为优秀, 三小时至十小时为一般, 三小时以下为较差				

(二) 模型构建

Bert 预训练模型在大量的无监督语料上训练出共同特征的语言表示, 预训练的特征也使得 Bert 的在获得更好的预测能力的同时提高模型泛化能力。另外, 本文在 Bert 的基础上做了微调以适应下游任务, 同时为了输出最后的评价等级, 在 Bert 的原有结构添加 softmax 线性层, 模型的结构如图 1 所示。具体地, Bert 采用 12 层的相同编码层, 在第一层是输入层, 其将经过标签归一化的数据作为输入序列 X, 随后 X 经过词嵌入模块进行词向量编码操作, 将输入序列从低维空间映射到多维空间, 每个输入的单词词向量由字编码、位置编码、分割编码相加得到向量表示 V, 进一步将输入向量 V 经过 Bert 的编码层进行特征提取得到输入序列的语言特征, 并经过多层编码层处理后将输出向量传入 softmax 线性层转化为概率输出并映射到不同的标签。

(三) 模型微调与优化

由于 Bert 是在大型无监督语料库中得来的, 在进行实验时需要根据具体任务进行微调。为此, 本模型主要从两个方面进行微调, 一方面采用增加一个特征提取模块, 实现对数据进行特征处理; 另一方面通过调整模型学习率、训练迭代次数、模型超参数等提高模型性能。

(四) 模型评价

为了实现模型评价, 本文采用准确性、敏感度、特异性进行性能评估。定义 TruePositive (TP)、TrueNegative (TN)、FalsePositive (FP)、FalseNegative (FN) 分别表示模型预测真阳性、真阴性、假阳性、假阴性数量, 则三种评价指标定义为:

(1) 准确性: 表示真阳性和真阴性样本数量占总测试样本的比例。

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

(2) 灵敏度: 表示真阳性样本占总阳性样本的比例。

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

(3) 特异性: 表示真阴性样本占总阴性样本的比例。

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

四、小结

在人工智能技术高速发展的当代, 线上教学的评价系统也可以依托现代技术进行完善。本文基于人工智能 Bert 模型, 将线上教学评价任务和自然语言处理传统任务结合, 通过对初始的教学数据预处理和对 Bert 模型的微调, 使改进模型能够作为线上教学可用的评价模型, 以供教师掌握学生的学习情况作出反馈。

参考文献:

[1] 魏伊非. 人工智能在教育领域的应用 [J]. 集成电路应用, 2019, 36 (02): 78-80.

[2] 陈国心. 基于自然语言处理的评教文本分类与分析 [J]. 西安航空学院学报, 2021, 39 (03): 91-96.

[3] 余同瑞, 金丹, 韩晓臻, 李家辉, 郁婷. 自然语言处理预训练模型的研究综述 [J]. 计算机工程与应用, 2020, 56 (23): 12-22.

[4] 晋百川, 杨鸿波, 胡大胆. 面向 XGBoost 的课程评价文本智能分类模型 [J]. 软件导刊, 2021, 20 (09): 14-17.

[5] 李琳, 董璐璐, 马洪超. 基于 Bert 的汉语作文自动评分研究 [J]. 中国考试, 2022 (05): 73-80.

[6] 张铭泉, 周辉, 曹锦纲. 基于注意力机制的双 Bert 有向情感文本分类研究 [J]. 智能系统学报, 2022, 17 (06): 1220-1227.

作者简介:

张国凯 (通讯作者), 男, 1991 年 9 月出生, 山东淄博人, 讲师, 博士, 硕士生导师, 研究领域为人工智能、大数据分析;

江浩玉, 女, 安徽黄山人, 硕士研究生, 研究领域为自然语言处理; 卢金典, 男, 山东莱芜人, 硕士研究生, 研究领域为大数据分析处理。