Advances in Education 教育进展, 2024, 14(12), 134-139
Published Online December 2024 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2024.14122244

Hans汉斯

AI辅助课程作业鉴别在高校教育中的应用研究

李　娜*，王小俊，彭敦陆

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2024年10月30日；录用日期：2024年11月28日；发布日期：2024年12月6日

摘　要

为应对AI辅助课程作业给高校教学秩序和学生培养带来的潜在风险，提出一种自动鉴别AI辅助作业的方法。在广泛收集AI生成、AI润色和人类撰写作业的基础上，采用对比学习技术深入挖掘有效区分AI辅助与人类撰写作业的文本特征，并基于这些特征构建一个高效准确的智能鉴别模型。在测试集上的准确率达到92.23%，实现了对不同类别作业的准确鉴别。研究成果不仅为有效的课程作业评估提供准确依据，更为维护公平、良好的教学秩序提供了有力支持。

关键词

AI辅助课程作业，智能鉴别，高校教育，对比学习

Research on the Application of AI-Assisted Course Assignment Identification in Higher Education

Na Li*, Xiaojun Wang, Dunlu Peng

School of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Oct. 30th, 2024; accepted: Nov. 28th, 2024; published: Dec. 6th, 2024

Abstract

A method for automatically identifying AI-generated course homework is proposed to address the potential risks that AI-generated homework brings to teaching order and student training. Based on the extensive collection of AI-generated, AI-polished, and human-written homework, contrastive learning technology is used to deeply explore the text features that can effectively distinguish between

*通讯作者。

文章引用: 李娜, 王小俊, 彭敦陆. AI 辅助课程作业鉴别在高校教育中的应用研究[J]. 教育进展, 2024, 14(12): 134-139. DOI: 10.12677/ae.2024.14122244

AI-assisted and human-written homework, and an efficient and accurate intelligent identification model is constructed based on these features. The accuracy on the test set reached 92.23%, achieving accurate identification of different types of homework. The research results not only provide an accurate basis for effective course assignment evaluation but also provide strong support for maintaining fair and good teaching order.

Keywords

AI-Assisted Course Assignments, Intelligent Identification, Higher Education, Contrastive Learning

1. 引言

近年来，语言大模型的快速发展推动了生成式人工智能(AI, Artificial Intelligence)的突破性进展。诸如 ChatGPT、GPT-4、文心一言、通义千问等大语言模型凭借其卓越的文本生成能力，受到广泛关注，并被应用于多个领域[1]-[3]。在教育领域，生成式 AI 的应用也取得了显著成效。从学生智能辅导到教师教学辅助[4] [5]，生成式 AI 技术正逐步改变着传统的教学模式。这些应用不仅提升了教学效率和质量，还为学生提供了更加便捷、高效和个性化的学习体验。教师可以更精准地掌握学生的学习情况，制定更加科学合理的教学计划；学生则可以根据自身的学习需求和能力水平，选择适合自己的学习资源和路径，实现个性化发展。

然而，随着生成式 AI 的日益普及，其潜藏的滥用风险亦不容忽视，可能对教学秩序和学生培养产生不良影响。一些学生可能利用先进的语言大模型轻松完成课程作业，违背学术诚信和公平竞争原则，扰乱正常的教学秩序。同时，这种情况也使得教师难以准确评估学生对课程知识的真实掌握程度。由于 AI 生成的文本具有高度的逼真性和连贯性，非常接近人类真实撰写的内容，教师很难仅凭人工判断学生作业是否由 AI 辅助生成。更为严重的是，如果这一现象持续蔓延，将会对学生的学习态度和价值观产生负面影响，导致他们忽视勤奋学习和独立思考的重要性，转而过度依赖 AI 工具来完成学习任务。因此，研究高效精准的自动鉴别 AI 辅助生成作业的方法，对于课程作业评估、维护教学秩序、塑造学生学习观念和价值观至关重要。

本研究提出了一种基于对比学习的自动鉴别 AI 辅助课程作业的方法。该方法通过对比学习技术学习 AI 辅助作业与人类撰写作业之间的微小特征差异，并构建智能鉴别模型，实现对学生作业真实性的高效、准确鉴别。这一方法不仅可以为教师评估课程作业成绩提供可靠依据，有助于维护学术诚信、保障教学质量，还能够为 AI 技术在教育领域的健康发展提供有力支持。

2. 研究方法与思路

本研究旨在设计一种自动鉴别 AI 辅助课程作业的方法，为课程作业评估提供可靠依据，维护教学秩序。为实现这一目标，我们首先收集 AI 辅助和人类撰写作业数据，建立数据集；随后，通过对比学习技术深入挖掘不同类别作业数据的特征表示，使不同类别作业数据的特征尽可能不同；最后，利用收集的数据集构建并测试自动鉴别模型，以识别学生作业是否为 AI 辅助生成。

2.1. 数据收集

为了构建有效的自动鉴别模型，首先需要建立一个全面且具有代表性的数据集。本研究主要收集三

种类别的作业数据：人类撰写作业、AI 生成作业和 AI 润色作业。

- 人类撰写作业：作业完全由学生真实撰写，体现了人类的写作习惯。为获取这类数据，我们收集了上海理工大学光电信息科学与工程、自动化、计算机科学与技术、数据科学与大数据技术等专业 2019 至 2021 级所有本科生完成的 10 门不同课程的作业。不同专业课程不同，作业形式不同；不同学生对作业题目的理解和撰写风格也存在个体差异。因此，这些数据能确保样本的广泛性和多样性。
- AI 生成作业：作业完全由 AI 生成，无任何人工干预。为了获取此类数据，基于作业题目和要求，通过调整作业题目的语序和表述，生成多种提示输入，并使用文心一言、通义千问、ChatGPT 等大语言模型生成大量作业文本。这些作业涵盖了不同专业的作业题目，而且由多种不同的语言模型生成，同样具备广泛性和多样性。
- AI 润色作业：在原有人类撰写作业的基础上，通过大语言模型进行润色。具体方法是将“请帮我润色以下内容：”与人类撰写的作业内容拼接作为输入，利用文心一言、通义千问、ChatGPT 等大语言模型对作业进行润色。这类作业既保留了人类撰写的特点，又融入了 AI 生成的痕迹，给模型鉴别带来挑战。

2.2. 文本特征学习

为了实现自动鉴别模型的学习，需要对课程作业文本进行特征编码，将其转换为特征向量。图 1 所示为本研究提出的基于对比学习的文本特征学习框架。如图所示，本研究首先利用先进的语言模型提取作业文本的初始特征，然后构建一个基于对比学习的文本特征学习网络来进一步学习不同类别作业之间的特征差异，获取更具有区分度的作业文本特征，为后续鉴别提供基础。

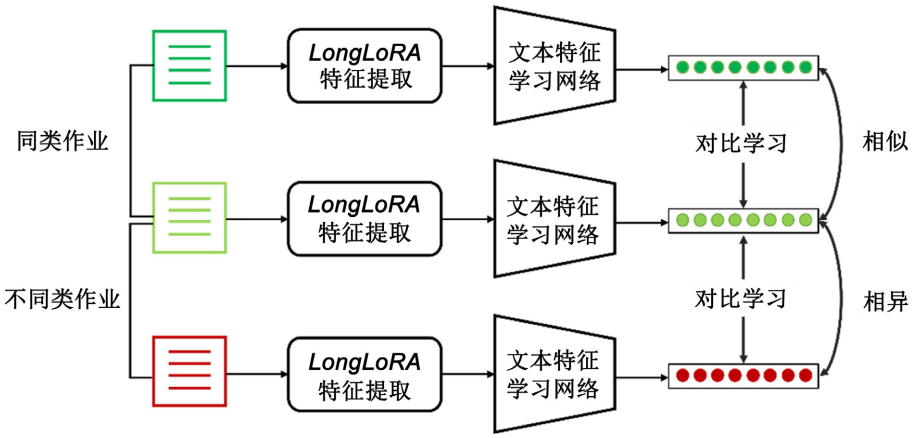

Figure 1. Framework of textual feature learning based on contrastive learning
图 1. 基于对比学习的文本特征学习框架

由于课程作业通常由大量字词组成，一般的神经网络模型(如循环神经网络、卷积神经网络)只支持短文本输入，难以满足长文本作业数据的处理需求。本研究采用 LongLoRA 模型[6]作为特征编码器，提取作业文本的初始特征。LongLoRA 模型是一个专为处理长文本而设计的开源语言模型，它继承了 Transformer 结构的强大性能，并进行了针对性的优化以适应长文本数据的处理需求。该模型能够捕捉作业文本中的长距离依赖关系，并生成高维、丰富的特征表示，这些特征表示包含了作业文本的词汇、语法、结构、语义及情感等多维度信息，可以挖掘作业文本的深层语义信息。本研究取 LongLoRA 模型最后一个隐藏层的输出作为文本的初始特征向量。

在提取作业文本初步特征的基础上，本研究构建一个由多个非线性神经网络层构成的文本特征学习网络，通过对比学习技术进一步学习具有高区分度的作业特征，为作业鉴别提供有力依据。特征学习网络的具体结构包括线性层、激活层和归一化层等，其中线性层可以对作业初始特征进行线性转换，激活层对线性层输出进行非线性转换，归一化层具有正则化效果，有助于防止过拟合。

对比学习是一种通过比较样本之间的相似性和差异性来学习有用特征表示的技术。本研究利用对比学习这一特性进一步学习作业文本的特征，使得同类作业的特征更相似，而不同类作业的特征更具有差异性和区分度。具体来说，给定某一作业样本，将其同类作业视为正样本，与之组成正样本对；其他类别作业视为负样本，组成负样本对。以对比损失(如 InfoNCE)为目标函数，将这些正负样本对输入到文本特征学习网络中进行训练，使网络学习到能够有效区分不同类别作业的特征表示。

2.3. 智能鉴别模型构建

AI 辅助课程作业的鉴别本质上可视为一种分类任务，即将输入作业划分为人类撰写作业、AI 生成作业和 AI 润色作业三种类别。为此，本研究在文本特征学习网络的基础上，添加一个多层感知机神经网络作为自动鉴别模型。该模型将文本特征学习网络学习到的作业特征转换为预测结果，实现作业分类。具体而言，多层感知机结构包括输入层、隐藏层和输出层，其中输入层接收文本特征学习网络输出的特征向量，输出层生成每个作业类别的预测概率。

对于自动鉴别模型的训练，本研究采用交叉熵损失作为优化目标。交叉熵损失是一种广泛应用于分类问题的损失函数，它能够衡量模型预测结果与实际类别之间的差异程度。通过最小化交叉熵损失函数值，模型能够不断调整其参数以更好地拟合训练数据，并提升在新数据上的泛化能力。

3. 实践评估

为了验证提出的基于对比学习的 AI 辅助课程作业智能鉴别方法的有效性，本研究进行了实践评估。首先，将收集到的三类作业数据随机打乱组合成一个完整的数据集，并标记所有作业的类别。随后，按照 6:2:2 的比例将数据集划分为训练集、验证集和测试集，分别用于模型的训练、验证和评估。

在实验中，文本特征学习网络的非线性网络层数为 3，激励层使用 ReLU 激活函数。自动鉴别模型隐藏层数量为 1，输出层采用 Softmax 作为激活函数，以适应多分类任务。本研究采用 Adam 来优化模型参数，以确保模型能够充分收敛。学习率设置为 0.01。在训练过程中，使用早停来防止过拟合。同时，利用验证集对模型进行了多次调参和性能评估工作，以选取最优的模型参数。

为全面评估方法的性能表现，本研究选择准确率、精确率、召回率和 F1 分数作为评价指标进行衡量。这些指标分别从不同角度衡量模型的整体鉴别能力，能够全面地反映模型的性能。因涉及人类撰写作业、AI 生成作业和 AI 润色三种课程作业类别，本研究计算宏精确率、宏召回率和宏 F1 分数，即精确率、召回率和 F1 分数在三种作业类别的宏观平均，以综合三个类别的指标。经过充分的实验验证和性能评估，本研究的方法在测试集上的宏精确率为0.9195，宏召回率为0.9233，宏F1分数为0.9204，准确率为92.23%，满足教师日常教学对学生作业真实性鉴别的需求，充分证明了对比学习在 AI 辅助课程作业鉴别中的有效性和可靠性。

4. AI 辅助文本鉴别方法对比分析

据作者所知，目前国内外针对 AI 辅助课程作业鉴别的研究较少，但对于 AI 辅助文本鉴别的研究却已有一定的积累[7]-[12]。现有的 AI 辅助文本鉴别方法大致可分为基于写作风格的方法、基于规则的方法、基于机器学习的方法、基于深度学习的方法等。

基于写作风格的方法通过分析文本的写作风格，包括词汇选择、句法结构、语法习惯等，与已知作者的写作风格进行比较，以判断文本是否由 AI 生成或润色。基于写作风格的方法能够个性化识别不同作者的写作风格，鉴别结果具有可解释性；但该方法需要已知作者的文本作为参考，不适用于对未知作者的文本鉴别。

基于规则的方法依靠专家预定义的规则集，包括语法结构、句子长度分布等特征，以区分人类撰写与 AI 辅助文本。基于规则的方法在简单的特定场景下具有一定的有效性，但其规则的设定往往需要高度的专业性，自然语言本身的多样性、复杂性和多变性使得规则难以被一一列举。此外，规则集可能过于僵硬，难以灵活地处理不同的文本风格和内容，导致较低的鉴别准确率。

基于机器学习的方法通过支持向量机、决策树、随机森林等算法，从大量已标记文本中学习鉴别模型，判断文本是否由 AI 生成。基于机器学习的方法与基于规则的方法相比，具有更强的灵活性和适应性，能够处理更加复杂和变化的文本。然而，基于机器学习的方法需要高质量的训练数据，且文本特征大多依赖人工定义的统计信息，如词频、句子长度、信息熵、词变异指数、词密度等，只能捕捉文本的浅层语义，降低鉴别准确率。

随着深度学习的发展，基于深度学习的方法不断涌现。利用深度神经网络强大的计算和学习能力，模型能够自动提取文本特征，捕捉文本中的依赖关系和复杂模式，自动鉴别 AI 辅助文本。基于深度学习的方法无需人工设计规则或特征，大大降低了前期的人工工作量。相比上述方法，基于深度学习的方法鉴别准确率最高。

本研究属于基于深度学习的方法。现有基于深度学习的方法通常使用卷积神经网络、循环神经网络等作为整体架构，依赖神经网络自身去学习 AI 辅助文本与人类撰写文本的特征，未考虑两者的高度相似性对鉴别造成的挑战和困难。本研究特别针对这一问题，利用对比学习捕捉 AI 辅助文本与人类撰写文本之间的细微差异，学习具有高度区分性的文本特征，从而提高鉴别准确率。此外，现有 AI 辅助文本鉴别方法通常只针对单一的文本形式，如文章、摘要等，本研究收集的课程作业则涉及不同的文本形式和内容，包括课程报告、程序代码、问答等，具有更高的文本多样性。

5. AI 辅助课程作业鉴别在高校教学中的影响

(一) 提升教学质量和效果

AI 辅助课程作业智能鉴别在教学中的应用，有助于提升教学质量和效果。通过及时鉴别和处理 AI 辅助课程作业，教师可以更加准确地了解学生的真实学习情况，从而调整教学内容和方法，使教学更加符合学生的实际需求。同时，还能够为教师提供更加全面的学生学习数据，有助于教师进行教学反思和改进。

(二) 引导学生树立正确的价值观

AI 辅助课程作业智能鉴别在教学中的应用，有助于引导学生树立正确的价值观。通过准确鉴别 AI 辅助课程作业，教师可以及时发现和纠正学生的错误行为，引导学生树立正确的学习观念，认识到勤奋学习和独立思考的重要性。同时，强化学生对于诚信、公正和责任感的价值观认同，为其未来的人生道路奠定坚实基础。

(三) 促进教学公平和诚信

AI 辅助课程作业智能鉴别在教学中的应用，有助于维护教学公平和诚信。通过准确鉴别 AI 辅助课程作业，教师可以对学生的失信行为进行及时处理，避免部分学生通过不正当手段获得高分，从而保障教学公平，维护良好的教学秩序。

6. 结束语

本研究提出了一种基于对比学习的 AI 辅助课程作业智能鉴别方法，并取得了一定的研究成果。通过收集多元化、高质量的数据集，设计合理的基于对比学习的文本特征学习网络，本研究成功学习到能够有效区分不同类别作业之间微小差异的特征，并构建自动鉴别模型实现 AI 辅助作业的准确鉴别。该方法的提出不仅有助于教师识别学生作业是否由 AI 辅助生成，引导学生树立正确的学习观念和学习态度，为公平有效的课程作业评估提供可靠依据，更为维护公平、良好的教学秩序提供了有力支持。

未来的研究可以进一步优化特征学习网络的结构，探索更多的特征表示技术，并在实际教学中进行实践与验证，以不断提升自动鉴别模型的准确性和适用性。同时，考虑到 AI 技术的快速发展，持续关注生成式 AI 的最新进展，并根据新的挑战不断调整鉴别策略，也是未来工作的重点。

基金项目

本文得到教育部产学合作协同育人项目(No. 220602510291241)、上海高校青年教师培养资助计划：《数据采集与集成技术》课程思政建设研究与实践(沪教委人[2023] 36 号)资助。

参考文献

[1] 王冰. 生成式 AI 技术在鞋类配色设计中的应用与分析[J]. 鞋类工艺与设计, 2024, 4(14): 198-200.

[2] 胡正坤, 谭覃. 生成式 AI 的大视听产业应用和未来展望——以商汤“日日新 SenseNova”大模型为例[J]. 视听界, 2024(4): 27-29+38.

[3] 邓元媛, 杨楠, 王子晴. 基于生成式 AI 的人工智能在建筑设计中的应用探究[J]. 智能建筑与智慧城市, 2024(7): 9-12.

[4] 王宇轩, 徐文浩, 于浩淼, 等. 生成式 AI 为 C 语言编程教学带来的挑战和机遇[J]. 计算机教育, 2024(8): 133-141+145.

[5] 徐倩芳. 生成式人工智能在小学信息科技教学中的应用探究[J]. 试题与研究, 2024(22): 121-123.

[6] Chen, Y., Qian, S., Tang, H., *et al*. (2024) LongLoRA: Efficient Fine-Tuning of Long-Context Large Language Models. *The* 12*th International Conference on Learning Representations*, Vienna Austria, 7-11 May 2014, 1-19.

[7] 北京理工大学. 一种基于写作风格的生成文本来源检测方法[P]. 中国专利, CN202410558437.0. 2024-08-06.

[8] 范志武, 姚金良. 基于深度金字塔卷积神经网络的 ChatGPT 生成文本检测方法[J]. 数据分析与知识发现, 2024, 8(7): 14-22.

[9] 戎蓉, 杨行, 韩叙, 等. 基于信息熵和 GBDT 算法的 AI 生成与人类撰写检测研究[J]. 信息技术与信息化, 2024(7): 186-189.

[10] 李帅彪. 基于 TPOC 识别与检测人工智能生成内容的研究与探讨[J]. 科技传播, 2024, 16(12): 18-22.

[11] 董腾飞, 杨频, 徐宇, 等. 基于事实和语义一致性的生成文本检测[J]. 四川大学学报(自然科学版), 2023, 60(4): 59-66.

[12] 王一博, 郭鑫, 刘智锋, 等. AI 生成与学者撰写中文论文摘要的检测与差异性比较研究[J]. 情报杂志, 2023, 42(9): 127-134.