

基于机器学习的大数据分析及其应用前景研究

张红姣

贵州师范大学

DOI:10.12238/acair.v2i4.10332

[摘要] 本文深入探讨了基于机器学习的大数据分析方法,结合两个相关项目研究,阐述了该方法在教育领域的应用。通过分析国内外研究现状,明确其创新点与特色,介绍技术路线及解决问题的方法,并对其应用前景进行了展望。重点关注了关键字识别技术、实时数据分析、学业预警功能以及个性化信息推送等方面,旨在为提升教学效率、推动教育信息化建设提供理论支持和实践参考。

[关键词] 机器学习; 大数据分析; 前景研究

中图分类号: TP181 **文献标识码:** A

Research on big data analysis based on machine learning and its application prospect

Hongjiao Zhang

Guizhou Normal University

[Abstract] This paper discusses in depth the big data analysis method based on machine learning, and describes the application of the method in the field of education by combining two related project studies. By analyzing the current status of research at home and abroad, it clarifies its innovations and features, introduces the technical route and problem-solving methods, and looks forward to its application prospects. Keyword recognition technology, real-time data analysis, academic warning function and personalized information pushing are focused on, aiming to provide theoretical support and practical reference for improving teaching efficiency and promoting education informatization.

[Key words] machine learning; big data analysis methods; prospect research

引言

随着信息技术的飞速发展,大数据时代已然来临。在教育领域,大量的学生学业相关数据不断产生,如何有效地利用这些数据成为提高教学效率和教育信息化建设的关键。机器学习作为一种强大的数据分析工具,为处理教育大数据提供了新的思路和方法。本文将结合两个相关项目研究,深入探讨基于机器学习的大数据分析方法及其在教育领域的应用前景。

1 国内外研究现状

1.1 国外研究现状

美国政府高度重视在线资源的制作及推送问题。涌现出三大知名在线平台: Coursera、Udacity和EdX,很多名校先后纷纷选择加入其中。著名的Data.gov网站使得资源的获取更加便捷。美国教育人士指出个性化学习资源推送是创造一个学习空间,构建个性化学习环境,且对于不同的学习者有不同的推送方式和内容,强调在资源推送中各要素不断变化,要不断完善。美国加州圣罗莎学区高中提出“套餐式教学”,充分体现了个性化资源推送理念。2014年,Facebook与SummitPublicSchools学校合作建立个性化学习平台,在大学生的个性化数据收集和推送方

面有所进展。2011年,加拿大不列颠哥伦比亚省颁布了“个性化学习”教育改革方案,强调利用信息技术构建“高技术”的学习环境促进资源的个性化推送。2016年加拿大著名远程教育和培训机构发布报告,指出学习具有移动性,列举了一批受欢迎的个性化学习新宠。

1.2 国内研究现状

在二十世纪九十年代,因特网迅速发展,为解决在线学习信息量过载和学习者个性化需求问题,个性化在线学习资源推送成为研究热点。2015年网易教育联合有道词典采集数千网友调查数据,总结了《中国在线教育趋势报告》。我国有各类在线教育类网站已达500余个。信息技术快速发展的时代,高校单独课堂教学已不能适应大学生个性化的知识获取需求,信息技术为个性化学习提供了充分的技术保障。

2 基于机器学习的大数据分析方法在教育领域的应用

2.1 项目一: 关键字识别与消息队列技术的信息推送系统研究目的:

解决教务信息下发、更新、推送不及时,接收与个人无关信

息过多问题。通过分析教务处提供的学生学业相关数据并识别关键事件,系统能够精准排序并推送课程信息、作业安排和考试提醒,减少冗余信息干扰,提高学生学习效率。为未修满学分的学生提供学业预警功能,帮助提高学业成绩并顺利毕业。

研究内容:

(1) 关键字识别技术的应用: 利用自然语言处理技术和关键字提取算法,从教务数据中识别和提取关键信息,为后续的推送和分析提供基础。

(2) 实时数据分析和事件识别: 建立实时数据分析系统,监测并分析教务数据流,及时识别与学生学习 and 教学管理相关的重要事件,快速、准确地推送相关信息。

(3) 学业预警功能的开发: 开发学业预警系统,帮助学生及时获取课程信息、选课建议以及学习规划,提供针对性的学业预警,调整学习策略。

创新点与项目特色:

(1) 创新点。关键字识别技术的应用: 采用自然语言处理技术和关键字提取算法,提高信息处理的准确性和效率。

实时数据分析和事件识别: 建立实时数据分析系统,及时识别重要事件,实现快速、准确的信息推送。

学业预警功能的开发: 结合实时数据流进行学生分类与预警,提供针对性的学业建议和预警。

(2) 特点。综合性: 整合关键字识别、实时数据分析和学业预警等技术,为学校提供全面的教学管理支持。

高效性: 采用先进技术和算法,保证信息处理和分析的高效性和准确性。

智能化: 通过深度学习、消息队列、关键字识别技术,实现智能化的信息推送和学业预警。

技术路线及拟解决问题:

(1) 确保关键字识别的准确性: 采用RKN文本识别技术,结合TF-IDF算法和TextRank算法提取关键词,解决信息推送不准确的问题。

(2) 实时数据处理与分析: 采用K-means聚类算法与分布式并行计算,实现对教务数据的实时数据分析,解决信息推送不及时问题。

(3) 学业预警功能: 训练模型对实时数据流进行分类,检测需要补考与学分不足的学生,发送提醒。

2.2 项目二: 基于CNN-Transformer模型的大学生个性化学习资源推送研究

创新点与项目特色:

(1) 实时数据分析与关键事件识别: 利用深度学习技术对教务数据进行实时分析,自动提取关键信息,提高教务管理效率和精准度。

(2) CNN-Transformer模型构建与应用: 结合卷积神经网络(CNN)和Transformer模型的优势,用于识别和分类教务系统数据以及预测时序数据,帮助学校优化资源分配。

(3) 个性化信息推送: 利用协同过滤、内容过滤及混合推荐

算法,结合CNN-Transformer的输出结果,设计个性化推荐系统。

技术路线及拟解决问题:

(1) 实时数据处理与分析: 处理大量实时数据流,确保系统具备足够的计算能力和效率,实现及时通知和个性化推送。

(2) 关键字识别准确性和泛化能力: 深度学习模型在关键字识别方面具备较高的准确性和泛化能力,确保系统稳定性和可靠性。

(3) 个性化推送算法设计: 设计并优化个性化推送算法,包括数据预处理、模型架构、模型训练、模型评估与上线等环节。

3 基于机器学习的大数据分析方法的创新点

3.1 多技术融合提升信息处理能力

在教育领域,基于机器学习的大数据分析方法通过多技术融合显著提升了信息处理能力。传统数据处理方法已无法满足日益增长的数据需求,而融合自然语言处理、关键字提取、深度学习、卷积神经网络(CNN)和Transformer模型等先进技术,为教育大数据分析开辟了新途径。

自然语言处理技术能够有效理解和分析大量文本数据,如学生作业反馈和教师评语,通过精确提取关键信息为后续分析奠定基础。关键字提取算法进一步提高了信息提取的准确性和效率。深度学习则能自动识别数据特征,深入挖掘学生学习行为与成绩中的潜在规律。CNN在图像数据处理方面表现优异,适用于分析学生作业图像和教学视频,而Transformer模型擅长时序数据分析,对学生学习进度和课程安排进行建模,从而预测未来学习趋势。

CNN-Transformer模型的结合更是为教育大数据分析带来了新的突破。该模型充分发挥了CNN和Transformer的优势,既能处理图像数据,又能够处理时序数据。在对教务系统数据的分析中,CNN-Transformer模型可以准确地识别和分类各种数据,例如学生的选课信息、考试成绩分布等。同时,该模型还为个性化推荐提供了更强大的支持。通过对学生的学习行为、兴趣爱好等数据的分析,CNN-Transformer模型可以为学生推荐更加符合其个性化需求的课程、活动或资源。

3.2 个性化推荐满足不同需求

个性化推荐是基于机器学习的大数据分析方法在教育领域的又一重要创新。通过协同过滤、内容过滤及混合推荐算法,结合深度学习模型的输出结果,可以为不同学生和教师提供个性化的课程、活动或资源推荐。

协同过滤算法基于用户之间的相似性进行推荐。在教育领域中,可以通过分析学生的学习行为、兴趣爱好等数据,找到具有相似学习特征的学生群体。然后,根据这些相似学生的选课情况、学习资源使用情况等,为目标学生推荐可能感兴趣的课程和资源。内容过滤算法则根据课程、资源的内容特征进行推荐。例如,通过分析课程的主题、难度、教学方法等特征,以及学生的学习目标、知识水平等,为学生推荐符合其需求的课程。混合推荐算法则结合了协同过滤和内容过滤的优点,能够更加准确地满足学生的个性化需求。

深度学习模型的输出结果为个性化推荐提供了更加丰富的信息。通过对学生的学习行为数据进行深度学习分析,可以得到学生的学习风格、知识掌握程度等特征。这些特征可以作为个性化推荐的重要依据,为学生推荐更加适合其学习需求的课程和资源。例如,对于学习风格偏向视觉型的学生,可以推荐更多包含图像、视频等教学资源的课程;对于知识掌握程度较高的学生,可以推荐更具挑战性的课程。

这种个性化推荐能够更好地满足学生的学习需求,提高学习效率。学生不再需要花费大量时间在众多的课程和资源中进行筛选,而是可以直接获得符合自己需求的推荐。同时,个性化推荐也为教师提供了更有针对性的教学支持。教师可以根据学生的推荐结果,了解学生的学习需求和兴趣爱好,调整教学内容和方法,提高教学质量。

3.3 学业预警功能助力学生发展

开发学业预警系统,结合实时数据分析和关键字识别技术,对未修满学分的学生进行分类预警,提供针对性的学业建议和预警。这有助于学生及时调整学习策略,提高学业成绩,顺利完成学业。

4 基于机器学习的大数据分析方法面临的挑战及解决方法

4.1 数据获取与隐私保护

获取学生的学习行为数据、成绩数据等数据源和特征存在一定的困难。一方面,可以通过下发调查问卷收集数据,联系学校教务处合法获取分析学生成绩信息等方式来解决数据源不足的问题。另一方面,在数据获取和使用过程中,要严格遵守相关法律法规,加强数据隐私保护,确保学生的个人信息安全。

4.2 资金支持与资源分配

缺乏资金支持难以推动课程的开发和实施。可以向教育主管部门申请专项资金,参加赛事获得奖金,寻求赞助等方式来解决资金问题。同时,要合理分配资源,确保资金和技术资源能够有效地用于项目的开发和实施。

4.3 技术准确性与泛化能力

深度学习模型在关键字识别方面需要具备较高的准确性和泛化能力,以应对新出现的情况。可以通过不断优化算法、增加训练数据、采用更先进的深度学习技术等方式来提高模型的准确性和泛化能力。同时,要加强对模型的评估和监测,及时发现并解决问题。

5 基于机器学习的大数据分析方法的应用前景

5.1 提高教学效率

通过精准的信息推送和个性化推荐,减少冗余信息干扰,帮助学生更好地掌握课程信息、作业安排和考试提醒,提高学习效率。同时,为教师提供更有针对性的教学支持,优化教学资源分配,提高教学质量。

5.2 推动教育信息化建设

基于机器学习的大数据分析方法为教育信息化建设提供了新的技术手段和方法。通过实时数据分析和关键事件识别,为学校教学管理提供更准确的决策依据,提高管理效率和便捷性。同时,个性化信息推送和学业预警功能也为学生提供了更好的学习支持,促进教育公平和个性化发展。

5.3 拓展教育应用领域

除了在教育教学领域的应用,基于机器学习的大数据分析方法还可以拓展到教育评估、教育政策制定等领域。通过对大量教育数据的分析,可以为教育评估提供更客观、准确的依据,为教育政策制定提供科学的参考。

6 结论

基于机器学习的大数据分析方法在教育领域具有广阔的应用前景。通过融合多种技术,实现精准的信息推送、个性化推荐和学业预警功能,能够提高教学效率,推动教育信息化建设,促进教育公平和个性化发展。然而,在应用过程中也面临着数据获取、资金支持、技术准确性等挑战,需要采取相应的解决方法。未来,随着技术的不断发展和完善,基于机器学习的大数据分析方法将在教育领域发挥更加重要的作用。

[参考文献]

- [1]杨继成.云计算平台下的智能汽车大数据分析与应用[J].汽车维修技师,2024,(18):17-18.
- [2]张舒.基于数据分析与机器学习的洪水灾害预测与风险评估[J].信息技术与信息化,2024,(08):189-194.
- [3]武俊岭.基于大数据分析的网络安全事件检测与响应[J].互联网周刊,2024,(15):38-41.
- [4]朱琳.基于大数据分析的大学科技园数字化服务体系建设研究[J].信息记录材料,2024,25(08):61-63.

作者简介:

张红姣(2004--),女,汉族,贵州遵义人,本科,研究方向:大数据挖掘,数据科学与大数据技术,卷积神经网络,机器学习。