

基于大模型的云学习智能推送服务研究与设计

柴亚辉

江西软件职业技术大学 江西南昌 330013

摘要: 本文研究了基于大模型的云学习智能推送服务,通过分析大模型技术、个性化推荐系统与云计算在教育中的应用,设计了一套包含数据采集、处理、大模型服务、推荐算法与用户交互的系统架构。系统采用深度学习模型如BERT、GPT系列,结合在线/离线部署策略,构建精细用户画像,运用基于内容、协同过滤及强化学习的推荐算法,并优化云平台资源调度与监控,以实现高效、个性化的学习资源推送,促进教育公平与质量提升。

关键词: 大模型; 个性化推荐; 云学习; 智能推送

引言

随着信息技术的飞速发展,特别是人工智能技术的广泛应用,教育领域正经历一场前所未有的数字化转型。教育技术不仅改变了知识传授的方式,更深刻影响了学习体验与效果评估的模式。其中,大语言模型LLM(以下简称“大模型”),如ChatGPT、Qwen、Gemini、Gemma等,作为结合了大数据、大算力和强算法的自动生产内容的革命性工具^[1],其强大的语言理解和生成能力,为教育领域带来了全新的教学方式和内容创新,自动化和自适应的数据收集、分析以及评估反馈^{[2][3]}能够协助解放教师,呈现多领域高发展布局,为个性化学习提供了前所未有的可能性。

一、大模型技术概览

大模型,也称为预训练模型或大规模语言模型,如BERT、GPT-3等,近年来在人工智能领域取得了突破性进展。这类模型通过在互联网文本等大规模数据集上的自我监督学习,掌握了丰富的语言结构和语境理解能力,展现出惊人的生成、理解和推理能力。在教育技术领域,大模型的应用涵盖了内容生成、自动答疑、学习路径规划等多个方面,为学习者提供更加丰富、个性化的学习体验。研究指出,通过持续的模型迭代与优化,大模型

的教育应用潜力巨大,但仍需探索模型压缩、适配教育领域特定知识的有效途径。

个性化学习推荐系统旨在根据学习者的特点和需求,智能推送定制化的学习资源和路径。早期系统主要依赖于基于内容和协同过滤^{[4][5]}的推荐算法,而近年来,结合机器学习尤其是深度学习的方法逐渐成为主流,能够更好地捕捉学习者行为的复杂性和动态性。研究显示,融合学习者历史行为、情境信息、情感状态的大数据驱动推荐系统,能显著提升推荐的准确性和满意度。然而,如何平衡推荐的新颖性与舒适度,以及保护学习者隐私,依然是该领域面临的挑战。

二、云计算在教育服务中应用

云计算以其灵活性、可扩展性和成本效益,在教育领域得到了广泛应用。云平台不仅支持了大规模在线教育课程的运行,还促进了教育资源的集中管理与高效分发。通过云计算,教育机构能够轻松部署和维护复杂的教学系统,实现资源按需分配,满足不同场景下的教学需求。同时,云计算也为数据分析和智能决策提供了强有力的基础设施支持。现有研究表明,虽然云计算在教育中的应用取得了显著成效,但在数据安全、访问速度、服务连续性等方面仍需不断优化。

基于大模型的云学习智能推送服务展现出巨大潜力,但仍然面临着一系列挑战。首要的是数据隐私保护问题,如何在利用大数据优化推荐的同时确保个人数据安全,是一大难题。其次,模型的解释性不足限制了教育者对推荐结果的信任与接受度。此外,跨学科、多模态学习资源的整合,以及适应不同学习风格和能力水平的个性化策略设计,也是当前研究中的空白。未来的研究需要

基金项目: 江西省教育厅科学技术项目“基于情境感知的云学习智能推送服务系统建设研究”(GJJ191489)

作者简介: 柴亚辉(1976-06—),男,汉族,河北保定,江西软件职业技术大学,讲师,博士研究生,研究方向:教育信息化。

聚焦于构建更加透明、安全、高效的学习推荐系统，同时深入探索大模型在教育领域的创新应用，以期实现更加公平、高质量的教育服务。

三、系统架构设计

基于大模型的云学习智能推送服务系统架构，旨在构建一个高效、灵活且可扩展的学习支持平台。该架构从底层到顶层分为数据采集、数据处理与存储、大模型服务、推荐算法、及用户交互五层，每层各司其职，协同工作，确保个性化学习内容的精准推送，如图-1所示。



图-1 基于大模型的云学习智能推送服务系统架构

(一) 数据采集层

数据采集层是系统的基础，负责从多个源头收集学习者的行为数据、教育资源元数据、用户反馈等信息。这包括但不限于学习管理系统（LMS）日志、社交媒体互动、论坛讨论记录、以及学习者的直接输入。数据涵盖文档、图片、视频、音频数据，以及结构/非结构化数据等。使用API接口、Web爬虫、传感器技术和第三方数据集成，确保数据的全面性和时效性。

(二) 数据处理与存储层

数据处理与存储层对原始数据进行清洗、转换和归一化处理，构建数据仓库和数据湖，为后续分析提供结构化和非结构化数据支持，数据支持分布式存储。此层运用大数据处理框架（如Apache Hadoop、Spark）和NoSQL数据库（例如MongoDB）来处理高并发、海量数据的存储与查询需求，保证数据处理的高效性和存储的灵活性。

(三) 大模型服务层

大模型服务层集成并优化预训练模型，如BERT、GPT系列，针对教育领域进行RAG、微调和适配等，也可以采取垂直类大模型形式，并提供高级别的自然语言理解和生成服务。此层设计支持在线/离线模型部署策

略，确保模型响应迅速且资源高效利用，同时通过API形式为上层推荐算法提供强大的计算能力。

(四) 推荐算法层

推荐算法层基于大模型输出和用户数据，采用多种策略生成个性化学习资源推荐。包括基于内容的推荐，通过分析资源内容与用户兴趣的相似性；协同过滤，利用用户间的交互数据发现潜在偏好；以及结合两者优点的混合推荐策略。此外，引入强化学习机制，使系统能根据用户反馈动态调整推荐策略，优化长期学习效果。

(五) 用户交互层

用户交互层设计友好直观的界面，使学习者能够轻松浏览、搜索和接收个性化推荐内容。同时，收集用户反馈，实现智能交互，形成闭环，不断优化用户体验。该层集成移动优先和响应式设计原则，确保跨设备的一致性与可用性。

四、关键技术与方法

构建基于大模型的云学习智能推送服务所涉及的关键技术与方法，涵盖大模型应用、用户画像构建、智能推荐算法的实施，以及云平台的优化与弹性扩展策略。

(一) 大模型应用

1. 大模型选择与适应性改造。在教育技术场景中，选择合适的大模型是基础。当前，如BERT、Transformer-XL、GPT-3等模型因其强大的语言理解与生成能力而备受青睐。而直接应用这些通用模型可能无法满足教育领域的特定需求，因此需对其进行适应性改造，以增强模型在教育情境下的表现力和准确性。

2. 在线/离线模型部署策略。考虑到模型的体积庞大和计算密集型特性，采取有效的部署策略至关重要。在线部署允许模型即时响应用户请求，适用于低延迟应用场景，但资源消耗较大。离线部署则用于预计算和批量处理，适合处理大规模数据分析任务。

(二) 用户画像构建

1. 数据来源与特征提取。用户画像的构建依赖于多维度的数据收集，包括学习行为数据（如浏览、点击、完成度）、社交互动数据、用户个人信息及反馈。特征提取过程需从这些复杂数据中提取关键信息，如学习偏好、兴趣点、学习节奏等，为个性化推荐提供基础。

2. 行为模式识别。通过机器学习技术，特别是序列模式挖掘和聚类算法，识别用户在学习过程中的行为模式。这有助于预测学习路径、识别潜在的知识缺口，并据此提供定制化资源。

3. 动态更新机制。用户画像具备实时更新能力，以反映学习者兴趣的变化和进步。定期重评估用户行为和偏好，结合新产生的数据动态调整用户模型，确保推荐的时效性和相关性。

（三）智能推荐算法

1. 基于内容的推荐。基于内容的推荐方法侧重于分析学习资源的内容特征与用户历史行为的相似性，如利用TF-IDF、Word2Vec等技术计算资源与用户兴趣向量的匹配度。

2. 协同过滤与混合推荐。协同过滤通过分析用户之间的相似性或资源的共同偏好来进行推荐。基于用户的协同过滤关注用户间的相似度，而基于物品的则聚焦于资源的关联性。混合推荐策略结合了基于内容和协同过滤的优点，以提高推荐的准确性和覆盖率。

3. 强化学习。通过引入强化学习，模型在每次推荐后根据用户反馈（如点击、评分、停留时间）进行策略调整，不断优化推荐策略以最大化长期学习效益。这种动态调整机制使系统更智能，更能适应个体差异。

（四）云平台优化与弹性扩展

1. 资源调度与负载均衡。在云环境中，有效利用资源调度工具（如Kubernetes、Docker Swarm）自动分配计算资源，确保高并发场景下服务的稳定运行。负载均衡技术（如Nginx、HAProxy）均衡分发请求，避免单点过载，提升系统响应速度和容错能力。

2. 自动化运维与监控。实施自动化运维流程，如CI/CD（持续集成/持续部署），确保快速迭代和高效故障恢复。集成监控工具（如Prometheus、ELK Stack）实时跟踪系统性能指标和异常情况，为快速决策提供数据支持，保障系统的稳定性和安全性。

五、结论与未来工作

基于大模型的云学习智能推送服务的研究与设计，从教育技术的最新进展出发，分析了云学习环境下智能推送服务的迫切需求，构建了一套包含数据采集、处理、

大模型服务、推荐算法和用户交互的系统架构。在此基础上，详尽阐述了关键技术与方法，包括大模型的适应性应用、用户画像的精细化构建、智能推荐算法的创新融合，以及云平台的优化策略，为实现高效、个性化的学习资源推送提供了理论与实践指导。

基于大模型的云学习智能推送服务仍有广阔的研究空间。随着人工智能技术的不断演进，探索更为先进、高效的大模型，及其在教育场景下的深度定制与优化，将是重要方向；加强用户隐私保护与数据安全机制，确保在提供个性化服务的同时，维护用户信息的安全与隐私；持续关注学习者反馈，结合情感分析与适应性学习理论，进一步提升推荐系统的智能化水平，使之成为促进终身学习、支持个性化成长的强大工具。不断深化技术与教育实践的融合，将是推动教育技术领域持续创新的关键。

参考文献

- [1] 焦建利. ChatGPT: 学校教育的朋友? 还是敌人? [J]. 现代教育技术, 2023, (4): 5-16
- [2] Zhu M X, Liu O L, Lee H S. The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing [J]. Computers & Education, 2020, 143: 1-43.
- [3] Sarsa S, Denny P, Hellas A, et al. Automatic generation of programming exercises and code explanations using large language models [A]. Proceedings of the 2022 ACM Conference on International Computing Education Research [C]. New York, NY, USA, 2022: 27-43.
- [4] 江水. 基于协同过滤技术推荐系统的探究 [J]. 计算机技术与发展, 2021 (11): 1-7
- [5] 李雪婷, 杨抒, 赛亚热·迪力曼提, 等. 融合内容与协同过滤的混合推荐算法应用研究 [J]. 计算机技术与发展, 2021 (10): 24-29.