Global Science Publishing

基于联邦学习的个性化学习资源推荐系统研究

韩姣，付瑞*，孙昊，李碧锦，张玉涛
潍坊科技学院信息科学与工程学院，山东潍坊

摘要：智能时代，开展学生个性化学习资源推荐需要多平台、多模态的学习数据协同。然而，数据孤岛、隐私合规与冷启动等问题并存，成为该类系统落地的关键障碍。联邦学习作为“数据不动、模型协同”的新型范式，具备不共享原始数据而实现跨域建模的优势，可在保障隐私的同时提升推荐效能。基于联邦学习的核心思路，本文构建了面向教育场景的“云—端”个性化学习资源推荐系统实现架构：云侧负责安全聚合与参数分发，端侧部署轻量化个性化模块以兼顾全局共享与本地适配。为推动该技术的教学落地，本文围绕横向联邦学习、纵向联邦学习与联邦迁移学习三类典型情境，系统阐述数据组织方式与训练流程，给出学校教务平台、在线课堂与数字图书馆等多方协同的实现路径。研究为在隐私保护前提下实现跨平台资源共享与冷启动缓解提供了可行方案，为个性化教育服务的推广应用与可持续发展提供参考。

关键词：联邦学习；个性化学习推荐；隐私保护；数据共享；学习推荐系统

Research on a Federated-Learning-Based Personalized Learning Resource Recommendation System

Jiao Han, Rui Fu*, Hao Sun, Bijin Li, Yutao Zhang
School of Information Science and Engineering, Weifang University of Science and Technology, Weifang, Shandong

Abstract: In the era of intelligent education, effective personalized learning resource recommendation requires collaboration across platforms and modalities. Yet data silos, privacy compliance, and cold start issues hinder deployment. Federated learning keeps data local, trains models collaboratively, enables cross domain modeling without sharing raw data, and improves recommendation performance while preserving privacy. Building on this idea, we design a cloud and edge architecture, the cloud performs secure aggregation and parameter distribution, and the edge hosts lightweight personalization to balance global sharing with local adaptation. We present three scenarios covering horizontal federated learning, vertical federated learning, and federated transfer learning, and describe the corresponding data organization and training workflows. These scenarios provide implementation paths for collaboration among school administration systems, online course platforms, and digital libraries. This approach supports privacy preserving cross platform resource sharing and mitigates cold starts, providing a basis for scalable and sustainable personalized educational services.

Keywords: Federated Learning; Personalized Learning-resource Recommendation; Privacy Protection; Data Sharing; Learning-resource Recommendation System

1 引言

在“互联网+”时代，MOOC、超星学习通等学习系统为学生提供了海量的网络学习资源，不仅极大地拓展了学习方式的多样性，也在一定程度上提升了学习效率。然而，资源的海量化与异构化也带来了新的挑战[1]。学生常常在信息过载中感到无所适从，难以高效发现和筛选出真正适合自身认知水平、学习节奏与兴趣方向的优质内容。尽管协同过滤[2]、嵌入表征[3]、序列建模[4]等个性化推荐技术在一定程度上优化了资源匹配的精准度，但它们大多依赖于集中式的数据存储与处理机制，不仅面临数据隐私泄露的风险，也受限于“数据孤岛”现象——不同教育平台间数据难以互通，导致推荐模型泛化能力不足，同时也难以有效应对新用户或新资源的“冷启动”问题。

为了解决这些问题，谷歌[5]在2016年提出联邦学习理论，为构建隐私安全、跨平台协作的推荐系统提供了全新思路。作为一种去中心化的机器学习范式，联邦学习使多个参与方能够在本地数据不离开原始环境的前提下，协同训练全局模型。训练过程中，各方仅上传加密模型参数或梯度，而非原始数据本身，从而在提升模型整体性能的同时，从机制上降低敏感信息泄露的风险，并有效打破机构间的数据壁垒。目前，联邦学习已在医学[6]、金融风控[7]、物联网[8]等多个对数据隐私要求较高的领域展现出广泛的应用潜力。基于此，本研究聚焦于联邦学习在学生个性化学习资源推荐系统中的具体应用，探索其在保护数据隐私、数据孤岛、缓解冷启动问题等方面的可行路径。目标在于构建一个以联邦学习为底层架构的跨平台推荐系统，实现“数据不动而模型流动”的协同训练机制。借助这一机制，不仅能够汇聚多源教育数据以提升推荐结果的准确性与多样性，还能严格恪守数据合规要求，以期为教育数据的安全利用与智慧教学的精准供给提供新的解决思路。

2 学生学习数据共享及安全问题与联邦学习

2.1 学生学习数据共享及安全问题

在“互联网+教育”的背景下，学习平台与校内教务系统持续汇聚课程视频、课件、题库、测评记录及学生行为日志，为个性化推荐提供了坚实的数据基础[9]。然而，数据隐私、数据孤岛与冷启动等问题仍然突出，限制了资源共享的质量，也削弱了推荐系统效果的持续提升。

2.1.1 数据隐私

在个性化学习资源推荐系统的构建过程中，各数据持有方往往掌握着大量学生的敏感隐私信息，一旦发生泄露，其后果将难以估量[10]。现实案例中，教育平台的安全漏洞屡见不鲜。例如Edmodo学习社区因系统漏洞遭黑客攻击，导致7700万用户账户数据被窃取，严重侵害了学生隐私；在线教育平台Chegg更因安全防护体系薄弱，多次发生数千万级用户信息泄露事件。这些案例警示我们，在当下可信安全环境尚未完全成熟的背景下，开展个性化学习资源推荐时，学生隐私数据在传输、存储及使用各环节均面临极高的泄露风险。

2.1.2 数据孤岛

在智能教育时代，数据已成为驱动个性化学习资源推荐系统的核心要素[11]。当前支撑推荐模型的数据呈现多模态特征，既包括课程视频、课件文档等结构化教学资源，也涵盖题库作业、测评成绩等量化学习数据，以及点击流、停留时长等非结构化行为日志。数据来源同样呈现跨平台特性，涉及校级教务系统的学籍与成绩数据、教育信息化平台的资源使用记录、网络学习空间平台的社交互动信息等。然而受制于数据天然的分布式存储特性与开放共享的安全风险，这些多源异构数据长期处于“数据孤岛”状态——不同系统间因数据标准不统一、隐私保护机制差异、利益分配矛盾等因素，难以构建安全可信的流通通道。这种数据割裂现象直接导致推荐系统面临双重困境：训练数据集的碎片化严重限制了模型对学习者多维度特征的捕捉能力；跨平台协同推荐因缺乏全局数据视图，难以实现“千人千面”的精准推荐，最终制约了智慧教育生态中数据价值的深度释放。

2.1.3 冷启动

个性化学习资源推荐离不开历史交互数据。然而在实际教育场景中，系统常面临三类典型的“冷启动”困境[12]：第一，新生入学阶段因缺乏历史学习轨迹导致用户画像空白；第二，新课程或者新资源上线初期因交互数据稀缺难以评估内容质量；第三，跨平台迁移时因数据标准差异造成有效信息断层。这种数据真空状态直接导致推荐模型陷入窘境，既无法通过协同过滤算法捕捉群体行为模式，也难以利用内容分析法提取资源特征向量，最终使得新用户获得的推荐结果呈现显著的随机性与低相关性。因此，系统对教育大数据的采集和分析不仅是个性化资源推荐服务的数据基础，还是解决系统“冷启动”问题的关键。

综上所述，智能时代的学生个性化学习资源推荐需统筹考虑数据隐私保护、跨平台数据孤岛以及冷启动带来的稀疏性等挑战。为此，应构建一个在合规前提下运行、能够有效融合多源数据并持续优化的个性化推荐系统，实现精准推荐与隐私安全的兼顾。

2.2 联邦学习

联邦学习是在不汇聚各参与方原始数据的前提下，通过参数级协作完成联合建模的一种分布式机器学习范式，如图1所示。其基本思想是让数据“留在本地、只带模型走”，以此在遵循数据主权与合规要求的同时提升模型在多域数据上的泛化能力。与集中式训练相比，联邦学习能够在多源异构、跨机构的数据环境中开展协同建模，减少数据出域带来的隐私与治理负担。在教育场景中，联邦学习尤为适合校内不同学院、年级以及校际平台之间的协作式建模，有助于在不打破既有数据边界的情况下提升个性化学习资源推荐任务的效果与可持续性。

2.2.1 联邦学习的系统模型与训练过程

联邦学习的系统模型通常由中央服务器与多个参与者共同构成。中央服务器一般由联邦任务的发起方或其委托的云平台担任，负责协调任务执行、维护全局模型，并完成参数的聚合与分发。参与者则为数据拥有方，例如学校、学院、在线教学平台或学生终端等，各自持有本地数据集,在不共享原始数据的前提下协同训练共享模型参数。

假设联邦学习过程共执行轮，其基本训练流程可概括为以下三个步骤：

步骤一：模型初始化。中央服务器发布全局模型，并将其分发给所有参与方。

步骤二：模型训练与更新。各参与方基于接收到的全局模型参数，利用本地数据集进行训练模型，得到本地模型更新，并将该更新上传至中央服务器。

步骤三：中央服务器收集所有参与方上传的模型参数，通过聚合算法(如FedAvg)，生成更新后的全局模型，并将其再次下发至各参与方，来开启下

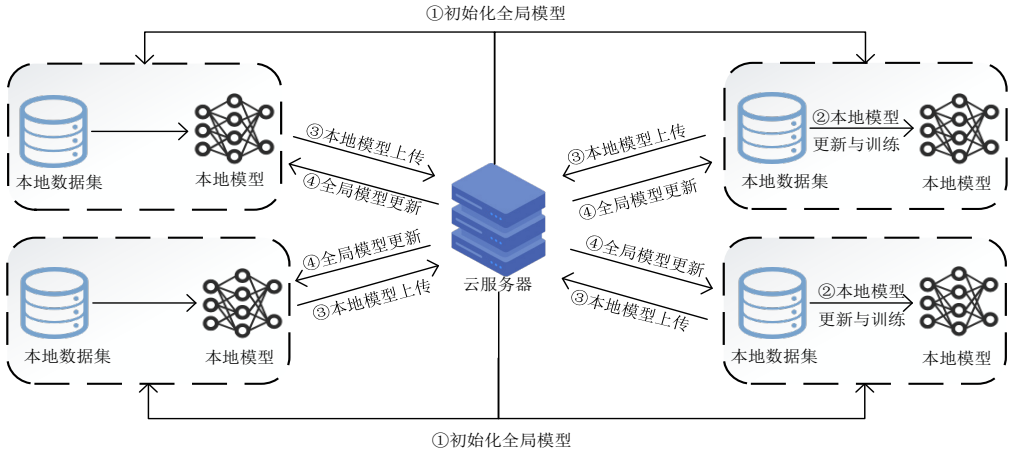

图1. 联邦学习框架

一轮的训练。

上述步骤二与步骤三循环执行，直至模型损失函数收敛或达到预设训练轮数，最终获得性能较优且具备泛化能力的全局模型。

2.2.2 联邦学习的分类

联邦学习根据参与方之间数据在特征空间与样本ID空间的重叠情况，主要分为三类：横向联邦学习、纵向联邦学习与联邦迁移学习[13]。

横向联邦学习适用于参与方的数据特征重叠较多、而样本重叠较少的场景。例如，多家医院希望联合构建疾病早筛模型，各家医院收集的电子病历具有相似的检查指标和结构（特征重叠高），但所服务的患者群体不同（样本重叠低）。通过横向联邦学习，各医院可在不交换原始病历数据的前提下，共同训练一个全局模型，从而提升早筛准确率，并满足医疗数据隐私与合规要求。

纵向联邦学习适用于参与方的样本重叠较多、而特征重叠较少的场景。例如，银行与电商平台希望联合构建信用风险评估模型。双方服务的用户群体高度重合（样本重叠高），但银行掌握用户的交易与信用记录，电商平台则拥有用户的消费行为与退换货记录（特征重叠低）。通过纵向联邦学习，双方能够在保护原始数据隐私的前提下，实现对同一用户多维度特征的“虚拟拼接”，从而训练出更全面的风控模型，提升欺诈识别与逾期预测的准确性。

联邦迁移学习适用于参与方之间数据特征与样本重叠均较少的场景。例如，在智能制造中，A工厂在金属件缺陷检测上积累了成熟的视觉检测模型，而B工厂生产的是材料与工艺完全不同的塑料件，两者数据分布差异大、特征不一致。通过联邦迁移学习，可以将A工厂的成熟模型作为“教师模型”，在联邦框架下辅助B工厂训练适用于自身生产条件的缺陷检测模型，实现知识迁移而无需直接共享数据。

2.2.3 联邦学习的隐私保护技术

联邦学习仅共享模型参数，而不共享本地数据集，虽然从根本上解决了数据拥有者的数据隐私问题。但是有研究者表明，攻击者仍可能基于所传递的参数推断出原始数据信息。因此为防止此类隐私泄露，可在联邦学习中引入隐私增强技术，作为数据安全的“加固层”。常见做法主要有三类：

一是基于差分隐私的隐私保护技术。该方案主要针对单个参与者被认出来的风险，核心做法是在参数上传前对本地更新先做裁剪，再按高斯机制加入少量噪声，使外部难以从共享参数中还原到某个具体参与者的信息。通过这一步的“轻微模糊”，把个人层面的重识别与成员推断风险降到很低，同时仍能完成协同训练。

二是基于加密技术的隐私保护技术。该方案主要针对服务器或者第三方能否看到某个参与方的单独参数问题。核心思想是是各参与方在本地使用同态加密等加密手段后再上传，服务器只能获得聚合结果，看不到任何单个参与者的明文更新。即使服务器或平台不完全可信，也能保障参与者的数据安全。也有研究开始探索将差分隐私与同态加密相结合，在加密前对参数添加噪声，形成双重保护机制。

三是基于可信执行环境的隐私保护技术。TEE通过硬件层面的隔离机制创建了安全的执行环境，如Intel SGX和ARM TrustZone。在联邦学习中，TEE可以用于构建可信的聚合服务，甚至承担部分训练任务。敏感的解密操作和参数处理过程在TEE的安全区域内完成，外部操作系统、云服务提供商乃至恶意攻击者均无法访问其中的明文数据，从流程层面保障了聚合与处理环节的机密性。

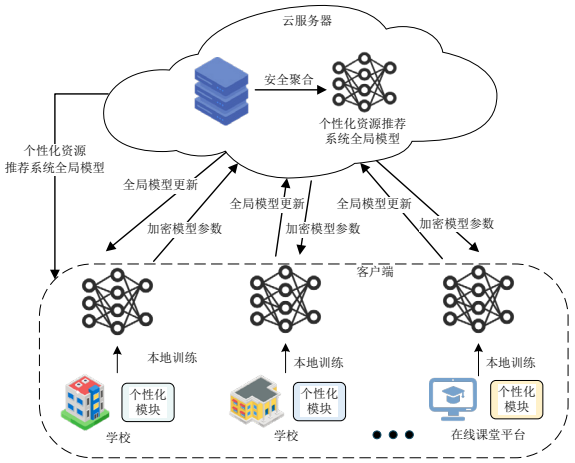

图2. 基于联邦学习的个性化学习资源推荐系统模型

3 基于联邦学习的个性化学习资源推荐系统

基于对学生学习数据共享与安全挑战、联邦学习技术及其隐私保护机制的综合分析，本研究设计了一种基于联邦学习的“云—端”两层个性化学习资源推荐系统模型，如图2所示。学生个性化学习资源具有来源多样、结构异构、动态演化等特征，其有效利用需依赖多方数据共享与协同建模。然而，传统集中式推荐系统难以在保障数据隐私与合规的前提下实现跨机构数据融合，成为当前智慧教育发展的主要瓶颈之一。针对上述问题，本研究构建的技术架构将系统划分为云侧与端侧两个逻辑层次，并基于“数据不动、模型流动”原则实现协同训练。其中，云侧被设定为可信第三方，负责全局模型的初始化、参数聚合与任务调度，不直接接触任何原始学习数据。端侧作为数据所有者与本地计算节点，承担模型训练与隐私保护等职责，仅向云侧上传经过加密或噪声处理的模型更新量。为适应教育场景中不同学校、平台或学生在数据分布与行为模式上的异质性，系统在端侧引入轻量级个性化适配模块，在全局共识与个体差异之间实现动态平衡。

在明确角色分工与架构设计的基础上，系统运行流程严格遵循“合规先行、安全协同”的原则。首先，依据国家《网络安全法》《个人信息保护法》及相关教育数据管理办法，在数据采集与交换前需获得学生或其监护人的明确授权，并完成数据脱敏与去标识化处理。在此基础上，云服务器发布初始化的全局推荐模型，各参与端（如学校服务器或可信学习终端）基于本地存储的学生行为数据、资源画像与上下文特征，对模型进行本地训练与更新。训练完成后，各参与方采用同态加密、差分隐私技术等隐私保护技术对模型参数进行加密或扰动处理，再将受保护的模型更新上传至云服务器。云服务器在接收到各端上传的加密参数后，在密文状态下执行安全聚合算法，生成新一代全局模型，并将聚合结果发回各参与方。各端对梯度进行解密后更新本地模型，进而开启下一轮训练。通过多次迭代，系统能够在不集中任何原始数据的前提下，逐步优化全局模型的推荐性能，最终形成一个既具备泛化能力又兼顾个体差异的个性化资源推荐联邦模型。

4 基于联邦学习的个性化学习资源系统的应用

依照上述论述可知，学生个性化学习资源推荐系统可在满足隐私约束的前提下，有效融合联邦学习的三种典型范式——横向联邦学习、纵向联邦学习与联邦迁移学习，以适应不同学校或平台之间在样本规模、特征空间及数据分布上的差异性，进而实现跨域协同建模。该系统能够在全程不暴露任何原始数据的基础上，完成学习资源的有效共享与安全防护。基于这一技术架构，本文接下来将围绕上述三类典型联邦学习场景，系统阐述推荐系统具体实现路径与应用中的关键要点，具体内容如下。

4.1 横向联邦学习的个性化学习资源推荐研究

学生个性化学习资源推荐通常涉及多类平台与场域，例如校内的教学管理平台、社会化的在线课程平台以及数字图书馆等。然而，由于商业边界与数据合规性的限制，这些平台之间的数据往往难以直接集中汇聚，从而对统一推荐模型的构建构成了挑战。尽管数据来源分散，但不同平台所记录的学生行为特征维度往往具有较高的重叠性。例如，学校A的教学管理平台、学校B的教学管理平台以及社会化在线课程平台C分别拥有不同学生的学习时长、完成率、作业测验得分、资源浏览与点击等学习数据，也就是说数据的特征维度存在较大的重叠，样本的重叠度较小。因此，这种情况下需要学校A、学校B以及在线课堂平台的所有学生样本数据作为本地数据集进行训练，进而开展基于横向联邦学习的学生个性化学习资源推荐模型构建，具体如图3所示。

设分别表示学校A、学校B、在线课程平台C的学生样本数据集，基于横向联邦学习的学生个性化学习资源推荐模型训练过程可以概括为：首先，云服务器初始化并下发联邦推荐全局模型。其次，各

数据拥有者分别用各自所属的学生数据集在本地独立地计算模型参数，随后使用加密技术对参数加密上传给云服务器。然后，云服务器实现安全聚合后下发全局模型参数给各端，各数据拥有者在本地解密并且更新模型。依次反复，直至模型达到收敛。最后，服务器对终轮本地更新进行聚合，输出学生个性化学习资源推荐模型。各平台将该模型部署于本地服务，结合各自资源库与学生画像生成推荐列表，实现跨平台、跨场域的资源精准推送与学习路径引导，同时在全流程中不暴露任何一方的原始数据与明文行为记录。

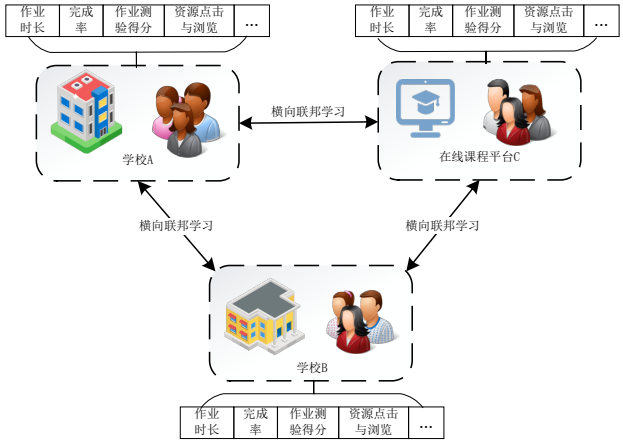

图3. 基于横向联邦学习的个性化学习资源推荐示例

4.2 纵向联邦学习的个性化学习资源推荐研究

学生个性化学习资源推荐中还面临另一类典型场景：同一学校或地区的同一批学生可能被多个平台共同覆盖，但各方所记录的数据特征重叠度较低。例如，学校教育管理平台A主要掌握学生的课程安排、作业测试、课堂评价与实验成绩等学习数据；在线课程平台B记录了学生在课外的学习时长、活跃度与学习轨迹等行为数据；而数字图书馆C则积累了学生的借阅、检索与阅读记录等数据。在此情况下，采用纵向联邦学习构建学生个性化学习资源推荐模型，如图4所示。

设分别为学校教育管理平台A、在线课堂平台B与数字图书馆C的数据集。基于纵向联邦学习的学生个性化学习资源推荐模型训练过程可以概括为：首先，采用加密的方式找出三方共同学生的匿名ID，在此基础上覆盖共同学生的三方特征并集，原始数据仍留在各自本地。随后，云服务器下发初始联邦推荐型和公共密钥。学校教育管理平台A、在线课堂平台B与数字图书馆C在本地分别利用对齐学生的本地特征进行训练，生成中间表示，并按选定的加密方案将受保护的中间计算结果进行交换与上报。然后将模型参数加密上传给云服务器，由云服务器安全聚合更新全局模型下发给各参与方，直至模型收敛。云端产出联邦推荐模型并分发至各平台本地部署，结合各自资源库与学生画像等生成最终的个性化推荐列表。

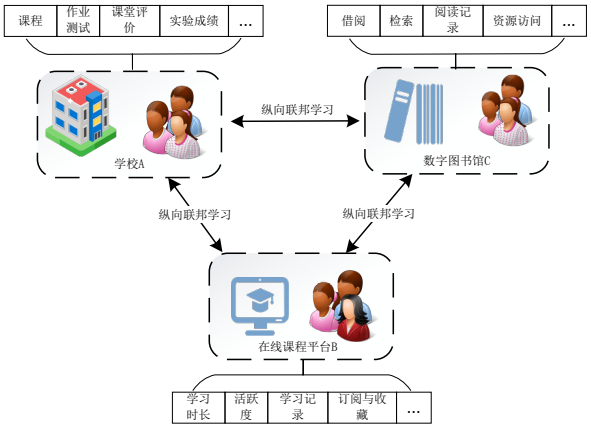

图4. 基于纵向联邦学习的个性化学习资源推荐示例

4.3 联邦迁移学习的个性化学习资源推荐研究

在新校区上线、特色课程新开或新专业等“冷启动”情形下，单独训练推荐模型常因数据稀缺而难以稳定。为此，可先在数据充足、业务相近的多方完成一次联邦训练，得到不含原始数据的可迁移参数；再将这些参数安全下发至数据稀缺的一方，用少量本地数据快速适配，从而在不共享明文数据的前提下，尽快获得可用的个性化推荐能力。以此为例：学校A与在线课程平台B拥有充足样本，而学校的新校区C样本较少，且与A、B在样本与特征上重叠度都不高。此时，开展基于联邦迁移学习的学生个性化学习资源推荐模型构建，如图5所示。

训练过程可概括如下：首先，云服务器将初始联邦推荐全局模型下发至学校教育管理平台A与在线课程平台B。两端基于各自本地数据对模型进

行独立训练，随后将加密后的模型参数或梯度上传至云端。云服务器对接收到的参数执行安全聚合操作，更新全局模型，并将更新结果再次下发给各参与方。该过程经过多轮迭代，直至模型收敛，最终获得具备较强泛化能力的通用模型参数。随后，云服务器将训练完成的可迁移参数包安全下发至新校区C。C端基于自身积累的少量本地数据，对通用模型进行小样本微调，使其快速适配本地场景，形成可用的个性化学习资源推荐模型。

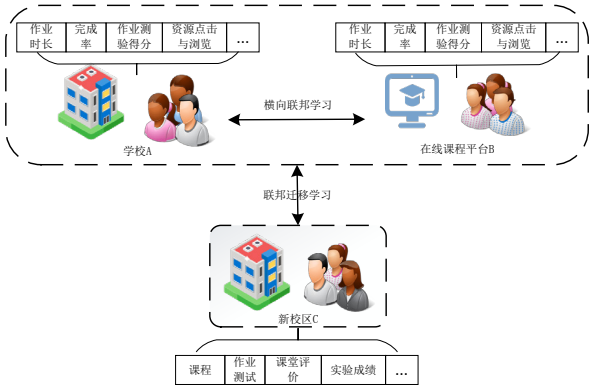

图5. 基于联邦迁移学习的个性化学习资源推荐示例

5 结论

在数字化学习迅速发展的背景下，尽管学习资源极为丰富，但数据隐私、数据孤岛与冷启动等问题制约了推荐系统的落地，学生难以从海量资源中高效选择，教学组织也受阻。联邦学习为此提供了可行路径：在不上传本地明文数据的前提下，仅以安全方式交换模型参数并进行聚合，实现跨域协作建模。基于此，本文构建了“云—端”两层的个性化学习资源推荐模型，并围绕横向、纵向与联邦迁移三类典型场景给出实现路径，兼顾资源共享与数据安全。未来工作将推进系统在真实教学环境中的部署与评测，重点验证其可用性与鲁棒性，并在推荐精度、冷启动适配与隐私保护强度等方面持续优化。

致谢

本文由以下基金项目资助：山东省自然科学基金（ZR2025QC649）；潍坊科技学院A类博士科研基金（KJRC2024006）；潍坊科技学院A类博士科研基金（KJRC2024014）。

参考文献

[1] Wang L. Collaborative filtering recommendation of music MOOC resources based on spark architectu re[J]. Computational Intelligence and Neuroscience, 2-022, 2022(1): 2117081.

[2] Zhai X, Wang Y, Liang L, et al. Personalized e-learning resource recommendation using multimodal-enhanced collaborative filtering[J]. Knowledge-Based Systems, 2025: 113605.

[3] 武杲昊, 王霞, 郝国生, 等. 融合知识图谱与高阶信息聚合机制的推荐模型[J]. Journal of Computer Engineering & Applications, 2025, 61(19).

[4] 彭梓航, 张全贵, 金海波, 等. 基于时间块动态图神经网络的序列推荐方法[J]. Application Research of Computers/ Jisuanji Yingyong Yanjiu, 2025, 42(8).

[5] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. PMLR, 2017: 1273-1282.

[6] 张畅, 李卫. 面向医疗健康领域的联邦学习综述: 应用, 挑战及未来发展方向[J]. 工程科学学报, 2025, 47(9): 1825-1840.

[7] Byrd D, Polychroniadou A. Differentially private secure multi-party computation for federated learning in financial applications[C]//Proceedings of the first ACM international conference on AI in finance. 2020: 1-9.

[8] Nguyen D C, Ding M, Pathirana P N, et al. Federated learning for internet of things: A comprehensive survey[J]. IEEE communications surveys & tutorials, 2021, 23(3): 1622-1658.

[9] 余胜泉, 李晓庆. 区域性教育大数据总体架构与应用模型[J]. 中国电化教育, 2019, 1: 18-27.

[10]刘生昊,吴国洋,邓贤君,张雨,鲁宏伟,何媛媛,杨天若.推荐系统与隐私保护研究综述[J].华中科技大学学报（自然科学版）,2023,51(2):1-9.

[11]张建楠.国内个性化学习研究热点及趋势分析——基于

CiteSpace 可视化分析[J]. Advances in Education, 2025, 15: 434.
[12]付文博,尹立杰.个性化推荐系统冷启动问题研究综述[J].新一代信息技术,2020,3(24):35-40.
[13] Yang Q, Liu Y, Chen T, et al. Federated machine learning: Concept and applications[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(2): 1-19.