

数据驱动的在线学习资源推荐框架研究

文 | 蒲菊华 王鋆珩 熊璋

摘 要:“互联网+”在教育领域的深入渗透和新冠肺炎疫情期间“停课不停学”的号召加速了在线学习平台的发展,但多样化学习资源的指数级增长容易给学习者(尤其是中小學生)造成知识迷航的困境。本文结合在线学习场景和需求,提出了一种融合依赖关系与序列表示的个性化在线学习资源推荐框架,充分利用多源数据,挖掘资源间固有关系和学习者动态兴趣变化,进行精准资源推荐,以消除信息壁垒,引导用户学习过程。

关键词:在线学习;学习资源推荐;个性化推荐;神经网络

作者简介:蒲菊华,北京航空航天大学计算机学院副教授;王鋆珩,北京航空航天大学计算机学院硕士研究生;熊璋,对外经济贸易大学信息学院院长,教授

中图分类号:G434

“互联网+”战略和教育信息化2.0进程的全面推进,使“互联网+教育”的全新教学理念得以广泛推广,中华人民共和国教育部印发的《教育信息化2.0行动计划》提出要逐步发展在线教学模式。2020年初爆发的新冠肺炎疫情为在线教育带来了新的机遇与挑战。据统计,2020年我国在线教育用户规模增长至4.23亿,占网民整体数量的46.8%。在政策支持、市场需求、技术革新等多方因素共同推动下,我国以中小學生为服务主体的K12在线教育快速发展,已成为当下的教育热点。

为迎合在线学习的市场需求,大量优质在线学习平台兴起并快速发展,因其可以以低成本而获得优质资源的契机而吸引了大量学习者,累积了包含各种粒度学习资源的海量数据,同时也不可避免地造成了信息过载问题,增加了学习者进行资源筛选的时间成本,使学习者难以获取满足个性化需求的匹配学习资源。统计数据显示,截至2020年底,我国上线慕课数量已超过3.4万门,而完课率却低于5%。

“信息过载”问题对K12用户群体的影响尤为严重,中小學生尤其是高年级学生课程多、日程满,绝大部分时间被线下教育占据,且中小學生对资源鉴别和自我认知能力相对较低。对海量在线资源进行筛选和对不匹配资源进行试错所耗费的时间与精力对K12在线教育目标用户群体而言是巨大的问题。

因此,在数据驱动的背景下,利用云平台、人工智能等前沿技术,构建针对中小學生在线学习需求的资源推荐系统是当前的重要发展方向。根据学习者的知识背景、个性特征和学习环境,向学习者提供个性化的学习资源推荐服务,帮助其进行高效的过滤和匹配,进而引导其学习;以此为提升中小學生自主学习意识和能力提供优质平台,全面提升用户在线学习的效率和体验。

一、在线学习资源推荐现状及挑战

(一) 在线学习资源推荐系统发展现状

学习资源推荐过程可以概括为三个主要环节:

数据收集、特征挖掘及推荐结果生成。在线学习资源推荐系统首先从学习平台中收集资源信息、学习者信息、学习者和资源间的交互日志（主要包括学习者的浏览、学习、答题行为）等多源数据，然后充分挖掘资源和学习者的特征，最后根据学习者和资源的匹配程度自适应地提供个性化推荐服务。

相较于发展相对成熟的电商推荐系统而言，学习资源推荐系统具有与其相似的模式和性质，故可以将学习资源推荐平台中的学习资源类比电商推荐系统中的物品，学习者类比为用户。因此，现有的在线学习推荐系统很多借鉴了电商平台中的商品推荐的主流思想和方法，迁移使用其中的主流推荐模型。

目前，学习资源推荐系统使用最为广泛的方法主要分为基于内容^{[1][2]}和基于协同过滤^{[3][4]}的推荐。简单来说，基于内容的推荐依据学习资源本身的属性计算资源间的相似度，然后根据学习者的学习历史，推荐和学习者曾经学习过的资源相似度最高的新资源；而基于协同过滤的推荐算法则利用统计数据和集体智慧，采取了“物以类聚，人以群分”的思想，大部分具有相似年龄层次、知识水平和兴趣爱好的学习者所需求的学习资源也是相似的，资源推荐系统可以从大量学习者及资源交互数据中学习某种隐性模式。

近年来，许多技术人员尝试对基于内容或协同过滤的模型进行改进^[5]，或融合先进技术提升课程推荐性能^[6]。但这些方法仍存在一定的局限性，难以充分挖掘资源推荐场景下的某些重要固有特征或难以满足更精细化的推荐需求。因此，我们有必要深度剖析在线学习资源推荐场景的特点，进而提出针对性的解决方案。

（二）在线学习资源推荐的两大主体特点

为充分挖掘在线学习资源推荐场景的独有特征并满足该场景下的一些精细化需求，我们针对在线学习资源推荐系统涉及的两大主体，即学习资源和

学习者，展开进一步深入分析，总结出各自重要特点。

对学习资源而言，学习资源不是相互独立的，不同学习资源之间可能存在复杂、隐性的依赖关系。以数学课程资源“基本初等函数”为例，其中包含多个视频资源“指数函数”“对数函数”“幂函数”等，视频资源间存在学习先后顺序要求，但资源间的这种先后关系通常较为隐性，难以直接获取。通过进一步分析发现，视频资源中包含有更细粒度的概念，如“指数幂”“对数运算”“换底公式”等，这些底层概念之间同样存在固有的先决条件依赖，而底层概念间的这些先后关系也应该被考虑和充分利用。

对学习资源而言，虽然类似的学习群体可能有大体一致的学习需求，但对于每一个学习者个体而言，其学习目的和知识背景各不相同，且短期学习兴趣可能发生动态变化。而这种短期行为所体现的学习兴趣可能对于下一个学习资源的选择更为重要。具体来说，学习者在平台上的学习行为呈现为按时间发展的顺序模式，我们将用户在较短的连续时间内交互行为序列称为一个会话。序列行为体现出三类特征：①学习者潜在学习偏好，即用户在一个会话内学习过的多个相似资源体现了其当前的潜在学习目的；②时序依赖特征，即学习者学习过的部分资源间存在严格的时间先后顺序，反映了某种隐性的依赖关系；③重复交互模式，即同一会话中学习者重复浏览或学习的资源体现出其较明确的意图和强兴趣。

因此，充分理解和利用序列行为中的此类特征，有助于理解学习者短期的兴趣变化，从而实现对其上下文进行更为精确建模表示，进而实现更为精准的学习资源推荐。

（三）在线学习资源推荐面临的挑战

通过对在线学习资源推荐系统两大主体的分析，我们发现将现有的基于内容和协同过滤的推荐方法

应用于在线学习资源推荐中会面临两大挑战。

第一, 现有推荐方法忽略了学习资源之间的依赖关系。如前所述, 学习资源间存在某些固有关系, 而非完全独立。现实情况下, 尤其是对于知识点体系较为庞大的基础课程而言, 许多资源之间存在先决条件依赖关系, 即学生需要按照某种顺序学习这些资源。这种依赖关系对于学习、组织、应用和生成知识至关重要。

第二, 现有推荐模型对学习者的个性化、动态学习偏好建模困难。在线学习平台面向的学习者通常具有较为明确的学习目的, 对推荐内容有较高的针对性和准确性要求; 同时每个学习者的知识背景、学习能力和学习目的各不相同, 且 K12 阶段教育涉及多方群体, 各方评价标准各异, 尤其是学生和家长对核心资源的质量和匹配度要求较高。基于内容和基于协同过滤的推荐模型建模是学习者长期的、静态的偏好, 忽略了其短期序列行为中的特征, 难以捕捉用户兴趣的漂移和变化。

二、以融合依赖关系和序列表示为思路构建学习资源推荐框架建设

针对在线学习资源推荐系统面临的上述挑战, 本文融合“依赖关系”和“序列表示”, 构建一种新的学习资源推荐框架, 充分考虑资源间的依赖关系和学习者的动态偏好, 以提升推荐性能。该框架主要由“资源依赖关系学习与挖掘模块”“基于图神经网络的序列表示模块”“依赖关系与序列表示的融合”三部分构成。

(一) 资源依赖关系学习与挖掘

资源依赖关系学习与挖掘模块是为了充分挖掘和利用资源间的关联与依赖关系, 以提升推荐效果和可解释性, 同时引导合理的学习路径。

对于在线学习平台而言, 学习资源众多, 资源之间以及资源所蕴含的概念之间的关系非常庞杂, 不可能依据人工的方式来建立所有的此类关系, 而

且学习资源间的有些依赖关系也是难以直接获取的。为此, 我们考虑将学习资源映射到更细的概念维度, 并通过学习概念图进一步推断资源间的依赖关系。

以课程资源依赖为例, 给定所有的课程, 课程间的真实关系标签体现了两个课程之间是否存在先后关系。理论上, 存在先后关系用 1 表示, 不存在先后关系用 0 表示。对于所有课程或学习资源间存在的海量复杂依赖关系, 我们采用数据挖掘和模型训练的方式, 将课程映射到通用概念空间, 并利用表示学习的技术方法得到每门课程的概念表示向量, 进一步学习两个概念对之间关系的有向权重, 通过训练好的模型自动推断所有课程间的依赖关系。

(二) 基于图神经网络的序列表示

基于图神经网络的序列表示是为了充分挖掘学习者的知识背景, 挖掘用户行为序列中的动态特征。

鉴于学习者偏好动态变化的特点, 我们考虑采用图神经网络对序列数据进行建模, 通过从学习者与学习资源交互(浏览、学习行为)序列中了解学习用户当前的学习兴趣, 并理解和建模学习者的兴趣偏好随时间的演变, 从而用表示向量来刻画学习者更精确、更动态意图, 提供定制、及时的序列化资源推荐。

第一, 根据学习者历史学习行为数据, 分别基于单个序列和全部序列构建会话图和全局图。第二, 采用门控图神经网络学习图中各个节点的向量表示。第三, 基于各节点重要程度动态融合节点表示当前学习者的最终偏好。第四, 通过计算学习者向量和资源向量间的相似性评分, 向学习者推荐评分最高的前若干个学习资源。

(三) 依赖关系与序列表示的融合

本文进一步融合上述资源依赖关系学习与挖掘模块和基于图神经网络的序列表示模块, 将学习到的依赖关系矩阵融合到序列推荐模型中, 以提升推荐性能。

该框架考虑两部分的融合, 如图 1 所示。

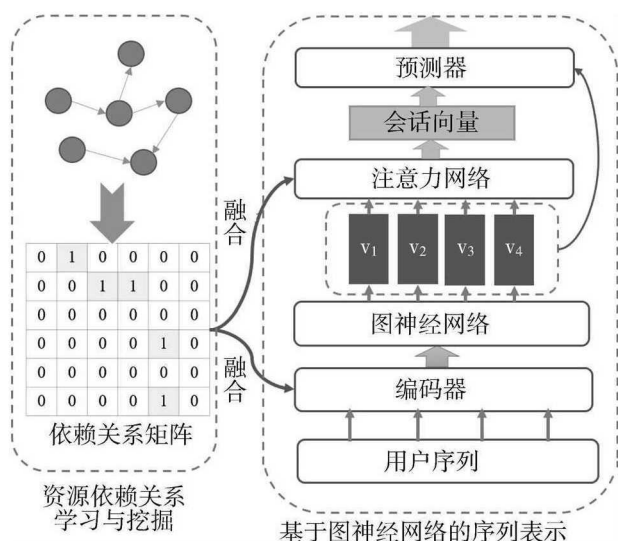


图1 融合神经网络与依赖关系的推荐框架

一方面,模型将资源依赖矩阵融合到编码器部分,在学习图节点表示时充分考虑资源本身与其他资源之间的依赖关系;另一方面,模型将资源依赖矩阵嵌入到注意力网络中,在为每个项目计算评分并生成推荐列表时,将资源间的先决条件关系纳入考虑。

综上,以融合依赖关系和序列表示为思路构建的学习资源推荐框架较好地解决了在线学习资源推荐面临的挑战。针对资源固有关系考虑不充分的问题,通过利用资源和概念间的映射关系进行先决条件学习,捕捉资源间依赖。针对学习者个性化学习目的及兴趣表示不准确的问题,从学习行为序列中挖掘学习者的潜在行为模式,并通过图神经网络建模会话间复杂转换来获取学习者的动态偏好变化,最终融合两个模块提供自适应资源推荐。

该框架可以推广于在线学习的各类学习资源,包括课程、视频、知识点、习题等。以习题推荐为例,该框架首先通过资源依赖关系学习与挖掘模块充分利用习题和知识间的映射关系,弥补传统推荐模型对资源间固有关系的忽视,发现习题间潜在结构或概念先后关系,保证推荐的习题具有典型代表性,且循序渐进,满足先决条件顺序要求。通过序

列表示模型还可以根据学生历史答题序列捕捉用户动态的知识状态和学习能力变化,以生成更契合中小学生学习能力水平的精准推荐。

三、总结

为了解决典型推荐方法应用于在线学习资源推荐面临的两大挑战,本文针对在线学习资源推荐的两大主体,即资源和学习者的特点,提出了一种融合依赖关系与序列表示的在线学习资源推荐框架。通过资源依赖关系的学习与挖掘模块和基于图神经网络的序列表示模块,分别解决了对资源和对学习者的特征学习不充分的问题,最后融合两部分进行自适应推荐,有效提升在线学习资源推荐的准确性。

本文系国家自然科学基金“面向在线智慧教育的自适应学习关键技术”(62177002)的研究成果

参考文献:

- [1] SHU J, SHEN X, LIU H, et al. A content-based recommendation algorithm for learning resources [J]. *Multimedia Systems*, 2018 (2): 163-173.
- [2] 黄冉. 基于内容和 word2vec 的慕课推荐算法研究 [D]. 济南: 山东师范大学, 2019.
- [3] EBESU T, SHEN B, FANG Y. Collaborative memory network for recommendation systems [C] //The 41st international ACM SIGIR conference on research & development in information retrieval, 2018: 515-524.
- [4] 石弘利, 王从瑜, 谢维奇. 基于协同过滤算法的教育平台课程推荐的研究 [J]. *电脑知识与技术*, 2021 (34): 19-22.
- [5] 张凡, 刘国营. 基于改进协同过滤算法的高校思政课程推荐系统研究 [J]. *微型电脑应用*, 2020 (9): 1-4.
- [6] 张晶晶. 融合图卷积网络与内容相似度计算的网络课程推荐研究 [D]. 武汉: 华中师范大学, 2021.