

文章编号: 2096-1618(2020)02-0151-07

基于注意力的深度协同在线学习资源推荐模型

冯金慧, 陶宏才

(西南交通大学信息科学与技术学院, 四川 成都 611756)

摘要:推荐可以看作是一个匹配问题,即为适当的用户匹配适当的项。针对学习平台和课程资源数量剧增以及在线资源分散使得课程推荐质量不佳等问题,将注意力机制和深度学习融入课程推荐问题中,提出一个基于注意力的深度协同在线学习资源推荐模型来为高阶课程集关系进行建模。该模型结合学习者信息和课程资源特征,学习用户和课程的隐线性特征和非线性特征,进行多模态特征拼合,融入注意力机制思想区分不同成对项目集对预测结果的贡献程度,以提升模型表示用户和课程的准确性,提高推荐性能。通过爬取慕课网(MOOC)上2014—2019年的学习数据进行实验,结果表明提出的模型在数据集 userlabel08rl 上多项评价指标要明显优于其它推荐算法。

关键词:深度学习;协同过滤;注意力机制;在线课程;推荐模型

中图分类号:TP391.3

文献标志码:A

doi:10.16836/j.cnki.jcuit.2020.02.005

0 引言

互联网高速发展,在线学习资源平台数量剧增,如国外的在线学习网站有 Udemy、Udacity、Coursera、Codecademy、MIT OpenCourseWare、EDX 等^[1];国内则有中国大学 MOOC、网易公开课、网易云课堂、学堂在线、腾讯课堂、尚学堂、哔哩哔哩、学习通、超星慕课等^[2]。虽然在线教育平台近十多年发展迅猛,课程资源大量积累、相关技术不断完善,相关的网站、课程资源、在线互动方式日渐丰富,但在线教育质量和在网上学习的学生效率总体低下^[3]。主要原因有:一是现有的在线平台增多,各个平台所提供的学习资源种类丰富、资源类型多样、课程资源难易程度不同,课程质量参差不齐,学生面临庞大的学习资源网站,各种纷繁复杂的在线课程学习资源,难以抉择;二是现有的在线学习资源系统普遍采取非常机械的方式将课堂从教室移到网上,缺乏对学生学习情况深入了解与分析,极少考虑学生的个性化需求,用户看到的学习内容几乎一致,极少考虑学习内容设计、页面之间的超文本链接跳转的合理性;三是每个学习平台的资源都是成千上万件,学习者想找合适的学习资源需要在各个平台目录中依次翻看,使得与学习者知识背景关联较强的课程资源查找变得烦琐复杂。现有的学习平台中缺少一个综合学习平台将主流的在线学习资源进行整合,然后在此基础上结合推荐技术将这些具有层次性、连贯性、系统性的

课程资源进行推荐^[4]。过度的自由选择让学习者苦不堪言,结合这一思想,本文通过以 MOOC 在线学习平台为基础,以各平台之间的各类型学习资料的相关数据为支撑,整合至统一学习平台,以方便学习者的资源查找和学习,同时利于课程提供者统一管理控制。其次,在此整合的在线学习资源平台上,提出一种基于注意力的深度协同过滤 (attention-based deep collaborative filtering, ADCF) 在线学习资源推荐模型,来为高阶课程集关系建模,通过分析每个学生的学习行为,结合学生个人的学习目标、意图、兴趣和结果,自适应地为学生推荐相关的学习材料。

本文拓展了广义矩阵分解模型,提出拓展广义矩阵算法 (expanded generalized matrix factorization, EGMF),该算法在协同过滤推荐上有更好的表现性能。下一步将融合 EGMF,并引入注意力机制,提出基于注意力的深度协同过滤 ADCF 推荐模型。该模型给出成对课程集关系建模的具体方法,不仅考虑了更多的课程和学习者信息,集成了 EGMF,还结合注意力机制思想,以此提高模型的推荐质量。在真实数据集上进行广泛的实验,验证了模型推荐的有效性。

1 相关理论

1.1 广义矩阵算法

协同过滤技术^[4] (collaborative filtering, CF) 是推荐系统中最为主流与经典的技术之一,传统的协同过

滤算法是基于矩阵分解的方法。矩阵分解(matrix factorization, MF)^[5], 为每个 User 和 Item 找到一个隐向量分解模型, 将效用矩阵分解为用户和项目的两个潜在表征矩阵, 并将它们反馈到网络中。令 \mathbf{p}_l 和 \mathbf{q}_c 分别表示用户(User)和物品(Item)的潜在向量; MF 评估相互作用 y_{lc} 作为 \mathbf{p}_l 和 \mathbf{q}_c 的内积, 则:

$$\hat{y}_{lc} = f(l, c | \mathbf{p}_l, \mathbf{q}_c) = \mathbf{p}_l^T \mathbf{q}_c = \sum_{k=1}^K \mathbf{p}_{lk}^T \mathbf{q}_{kc} \quad (1)$$

基于 MF 矩阵的思想, 广义矩阵分解(Generalized matrix factorization, GMF)^[6]是矩阵分解的一般泛化模型, 详细模型定义为

$$\begin{cases} \varphi_1(\mathbf{p}_l, \mathbf{q}_c) = \mathbf{p}_l \odot \mathbf{q}_c \\ \mathbf{p}_l = \mathbf{p}^T \mathbf{v}_l^l \\ \mathbf{q}_c = \mathbf{q}^T \mathbf{v}_c^c \\ y_{lc} = \alpha_{\text{out}}(\mathbf{W}^T(\mathbf{p}_l \odot \hat{\mathbf{q}}_c)) \end{cases} \quad (2)$$

式中 \odot 表示向量的逐元素乘积, α_{out} 表示激活函数, \mathbf{W}^T 表示连接权重。由式(2)可以看出, 当 α_{out} 为恒等函数且 \mathbf{W}^T 连接权重全部为 1 时, 恰好还原为传统矩阵分解^[7]。

1.2 多层感知器算法

多层感知器(multi-level perceptron, MLP)算法通过连接多个特征值, 经过线性和非线性组合, 最终达到可以判断是属于哪种类别的目标^[7]。MLP 模型^[5]定义为

$$\begin{cases} \mathbf{z}_1 = \varphi_1(\mathbf{p}_l, \mathbf{q}_c) = \begin{bmatrix} p_l \\ q_c \end{bmatrix} \\ \varphi_2(\mathbf{z}_1) = \alpha_2(\mathbf{W}_2^T \mathbf{z}_1 + \mathbf{b}_2), \\ \dots \\ \varphi_L(\mathbf{z}_{L-1}) = \alpha_L(\mathbf{W}_L^T \mathbf{z}_{L-1} + \mathbf{b}_L) \\ y_{lc} = \sigma(\mathbf{z}^T \varphi_L(\hat{\mathbf{z}}_{L-1})) \end{cases} \quad (3)$$

\mathbf{W}_x^T 表示第 x 层感知器的权重矩阵, \mathbf{b}_x 为第 x 层感知器的偏置, α_L 表示第 L 层的激活函数。

1.3 注意力机制

深度学习中的注意力机制^[8-12]主要是附着在编/解码器框架内^[7], 追本溯源, 注意力机制其实是一种通用的思想模型, 让一个神经网络只关注其输入的部分信息, 能够选择特定的输入, 具体模型为

$$\begin{cases} a = f_\varphi(\mathbf{X}) \\ \mathbf{Z}_a = a \odot \mathbf{Z} \\ \text{st: } a \in [0, 1] \end{cases} \quad (4)$$

式中, f_φ 是一个注意力机制网络, 生成 a , 将 a 与输入 \mathbf{X} 的特征向量 \mathbf{Z} 相乘。当 a 的取值范围为 $(0, 1)$ 时, 为软注意力机制; 当 a 只取 0 或 1 时, 为硬注意力机制。注意力网络可以拟合更复杂的函数模型。

1.4 深度神经网络

深度神经网络^[8](deep neural network), 简言之就是建立 m 个隐藏层, 按顺序建立输入层与隐藏层的联结, 最后建立隐藏层与输出层的联结, 为每个隐藏层的每个节点选择激活函数, 求解每个联结的权重和每个节点的偏置值, 其训练本质是通过反向传播和梯度下降相结合来求解。已知 y_1, y_2, \dots, y_n 和 x_1, x_2, \dots, x_n , 求解每个联结权值和每个神经元的偏差值, 用随机数初始化每个联结权值。然后, 通过神经网络计算出来预测值与真实的 y 值比较。如果这个值相差比较大, 则修改当前层的联结权重; 当发现这个值相差不大时, 则修改更低一层的权重。重复这个步骤, 逐步传递到第一层权值^[9]。

1.5 深度协同过滤模型

深度神经协同过滤模型(NeuMF)由 Xiangn-an He 等^[7]提出, NeuMF 模型由多层感知机(MLP)和广义矩阵分解(GMF)模型 2 部分组成。GMF 学习线性特征, MLP 学习非线性交互函数, 由此使得 NeuMF 模型同时具有线性和非线性建模能力, 泛化能力增强。NeuMF 架构主要由 MLP 和 GMF 两部分构成, 输入层输入的是 User_id 和 Item_id 的 one-hot 编码; 接着, 传入嵌入层进行学习表示相应的用户和物品特征; 然后, 将两部分学习到的特征作为深度神经网络的输入进行模型训练; 最后, 神经网络输出经过一个激活函数来获取预测值。NeuMF 模型为

$$\begin{cases} \boldsymbol{\varphi}^{GMF} = \mathbf{p}_u^{GMF} \odot \mathbf{q}_i^{GMF} \\ \boldsymbol{\varphi}^{MLP} = \alpha_L(\mathbf{W}_L^T(\alpha_{L-1}(\dots \alpha_2(\mathbf{W}_2^T \begin{bmatrix} \mathbf{p}_u^{MLP} \\ \mathbf{q}_i^{MLP} \end{bmatrix} + \mathbf{b}_2) \dots) + \mathbf{b}_L)) \\ \hat{y} = \sigma(\mathbf{h}^T \begin{bmatrix} \boldsymbol{\varphi}^{GMF} \\ \boldsymbol{\varphi}^{MLP} \end{bmatrix}) \end{cases} \quad (5)$$

式中, \mathbf{p}_u^{GMF} 和 \mathbf{p}_u^{MLP} , \mathbf{q}_i^{EGMF} 和 \mathbf{q}_i^{MLP} 分别代表用户和项目的特征向量, \mathbf{W}_x^T 表示第 x 层感知器的权重矩阵, \mathbf{b}_x 为第 x 层感知器的偏置, α_L 表示第 L 层的激活函数, \mathbf{h}^T 为权重系数矩阵。

2 一种新的基于注意力深度协同过滤模型

2.1 问题的提出

由前面对 NeuMF 模型的分析,发现 NeuMF 模型有几个可以改进之处。其一,输入部分仅以用户 ID 和项目 ID 的 One-hot 编码稀疏向量进行 Embedding 特征映射,考虑的特征较为单一,模型生成用户和项目表示时仅依靠 user_id 和 item_id 进行约束,没有考虑用户和物品的其他特征,忽略了用户和项目的本身特点对推荐结果的影响。其二,NeuMF 模型主要由 GMF 和 MLP 两部分组成,GMF 是 MF 的一种通用泛化模型,基本内核与矩阵分解等同,矩阵分解可能会有负数,在推荐中负数没有实际意义,也不利于推荐解释;GMF 用权重系数 \mathbf{W}^T 直接乘以用户和项目的关联性 $\mathbf{p}_i \odot \mathbf{q}_c$,可能会使对角线部分乘积较大,导致推荐结果倾向于只推荐物品本身。其三,NeuMF 模型认为所有与用户交互过的物品对预测用户和目标物品的贡献程度是等同的,这种认知方式存在一定的不足。比如,一个偏好

喜剧片的用户,偶尔会看恐怖、动作等其他类型的电影。在这种情况下用户历史交互项目中,喜剧类的电影相对于其他类型电影来说对预测目标电影的贡献程度应该更大。

据此提出基于注意力的深度协同过滤推荐模型 ADCF,并将其用于在线学习资源推荐中。ADCF 模型是一种引入了注意力机制思想,集成本文提出的 EGMF 算法,考虑用户和项目的其他特征信息的集成算法框架。

2.2 新模型的构建

ADCF 模型架构如图 1 所示,其与 NeuMF 模型架构类似,但结构上有三点不同:一是 NeuMF 模型输入层部分只对用户 ID 和物品 ID 进行建模,而 ADCF 模型输入层是基于学习者和课程以及其他附属信息组合的多维 one-hot-encoding 编码;二是 ADCF 嵌入层使用的是对 GMF 拓展后的 EGMF,而不是 GMF;三是考虑到学习者和不同的交互课程对模型的最终预测起不同程度的作用,引入了注意力机制,多模态融合层输出作为注意力机制的输入,利用注意力机制思想提取高阶交互信息,捕获各维度之间内在联系。

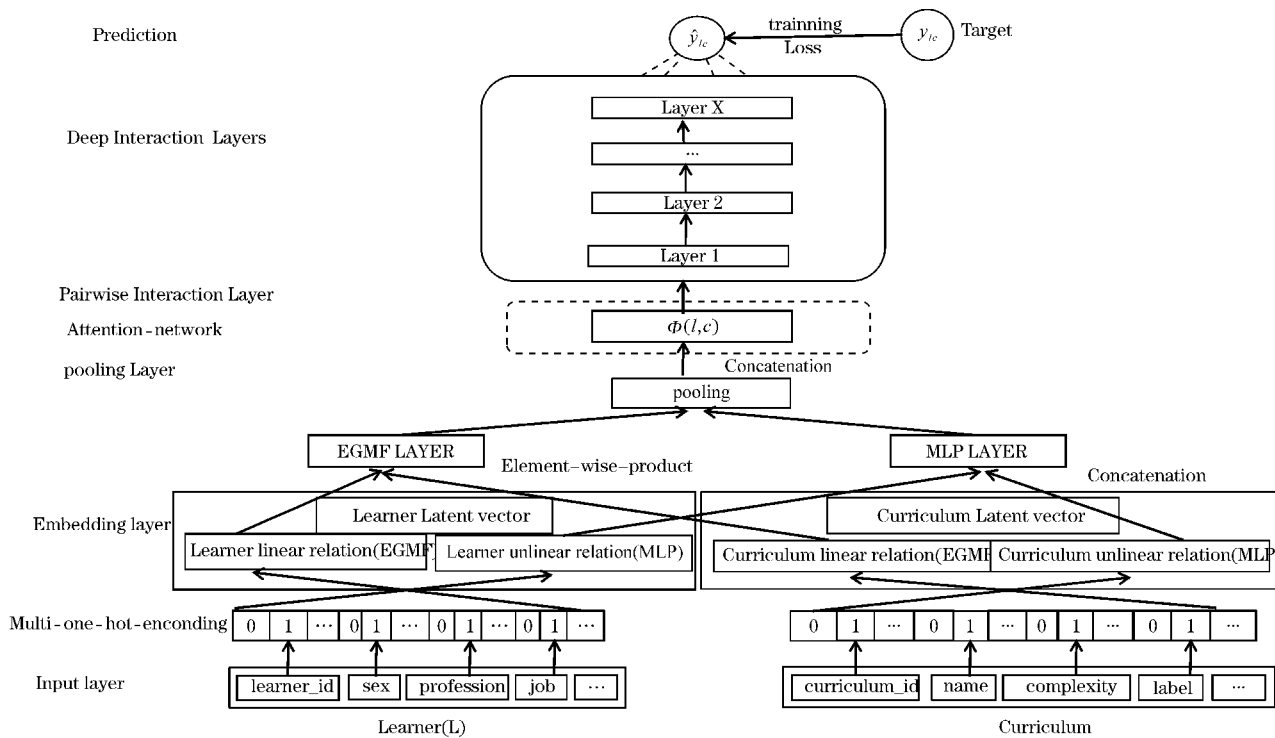


图 1 ADCF 模型架构

由图 1 可以看出,ADCF 模型分为 EGMF 和 MLP 两部分,所获得的嵌入向量可以看作是学习者的潜在向量 $\mathbf{p}_l, \mathbf{q}_c$ 表示基于辅助信息的课程资源的潜在向

量。嵌入层使用的是独立的 EGMF 和 MLP 嵌入层。左侧是 EGMF 模型分别对学习者和课程进行线性特征建模,右侧是 MLP 分别对学习者和课程进行非线性

关系的提取。接下来, pooling 层分别对 EGMF 和 MLP 的输出进行处理, 调整特征向量为同一规模大小, 便于下一步进行多模态特征融合层的特征拼接。多模态融合层输出作为注意力机制的输入, 根据学习者特征向量 \mathbf{p}_l 和学习者历史交互课程 \mathbf{q}_c 的交互特征计算出一个权重, 以区分不同的历史交互课程在预测目标课程时不同的贡献程度。注意力机制的输出作为神经网络的输入, 经过多次迭代训练, 最后经过一个激活函数得到的预测值, 进行推荐。

2.3 拓展广义矩阵模型

广义矩阵模型中, 矩阵分解成两个矩阵后可能产生负数, 而负数在用于推荐过程中没有实际意义, 也不利于推荐解释。为将广义矩阵模型更好地应用于推荐中, 本文提出了 EGMF, 其基本思想是, 原始效用矩阵 \mathbf{R} 被近似分解为两个低秩矩阵 $\mathbf{R} = \mathbf{P}^T \mathbf{Q}$ 相乘的形式, 其中 \mathbf{P} 、 \mathbf{Q} 两个元素需要满足非负约束条件即 $\mathbf{P} \geq 0$ 且 $\mathbf{Q} \geq 0$ 。同时, 为了防止过拟合, 加入正则项的乘性迭代, 则:

$$\begin{aligned} \mathbf{p}_{l,k} &= \mathbf{p}_{l,k} \cdot \frac{\sum_{i \in I_l} \mathbf{q}_{k,i} \cdot r_{l,c}}{|\mathbf{I}_l| \lambda_q \mathbf{q}_{l,k} + \sum_{i \in I_l} \mathbf{q}_{l,k} \cdot r_{l,c}^*} \\ \mathbf{q}_{k,c} &= \mathbf{q}_{k,c} \cdot \frac{\sum_{i \in I_l} \mathbf{p}_{l,k} \cdot r_{l,c}}{|\mathbf{U}_c| \lambda_q \mathbf{q}_{k,c} + \sum_{i \in I_l} \mathbf{p}_{l,k} \cdot r_{l,c}^*} \end{aligned} \quad (6)$$

式中, \mathbf{I}_l 为评分不为 0 的课程集合; \mathbf{U}_c 为评分不为 0 学习者集合; $r_{l,c}$ 为学习者 (learner) 对课程 (curriculum) 实际评分; $r_{l,c}^*$ 为预测的学习者对课程的评分, 其可以由初始化的学习者隐因子矩阵 $\mathbf{p}_{l,k}$ 和课程隐因子矩阵 $\mathbf{q}_{k,c}$ 计算得到。

定义 EGMF 算法模型为

$$\begin{cases} \boldsymbol{\varphi}_l(\mathbf{p}_l, \mathbf{q}_c) = \mathbf{p}_l \odot \mathbf{q}_c \\ \mathbf{p}_l = \mathbf{p}^T \mathbf{v}_l^l & \mathbf{p} \geq 0 \ \& \ \mathbf{v}_l^l \geq 0 \\ \mathbf{q}_c = \mathbf{q}^T \mathbf{v}_c^c & \mathbf{q} \geq 0 \ \& \ \mathbf{v}_c^c \geq 0 \\ \hat{y}_{lc} = \partial_{out}(\mathbf{W}^T(\mathbf{p}_l \odot \mathbf{q}_c)) \\ \text{st: } \mathbf{W} \geq 0 \ \& \ \text{diag}(\mathbf{W}) = 0 \end{cases} \quad (7)$$

由式(7)可以看出, EGMF 在 GMF 上进行了扩展。EGMF 模型多加了两个约束条件: 一是 \mathbf{W} (权重系数) 非负, 这样学习者和课程之间只有正相关关系, 利于推荐解释; 二是 \mathbf{W} 的对角线特征向量限制为 0, 与 $\mathbf{p}_l \odot \mathbf{q}_c$ 相乘之后, 为减小误差函数, 不会倾向于只推荐它自己, 以达到保证一个已知的行为得分不会用于预测成为自己的效果。神经网络的联结权重可以看作是对学习者和课程资源的线性加权, 非线性的激活函数可以捕捉传统 MF 不能学习的非线性特征。

2.4 ADCF 新模型的算法描述

ADCF 模型的数学表达为

$$\begin{cases} \mathbf{p}_l^{EGMF} = (\mathbf{l}_0^{EGMF} \oplus \mathbf{l}_1^{EGMF} \oplus \mathbf{l}_2^{EGMF} \oplus \mathbf{l}_3^{EGMF}) \\ \mathbf{p}_c^{EGMF} = (\mathbf{c}_0^{EGMF} \oplus \mathbf{c}_1^{EGMF} \oplus \mathbf{c}_2^{EGMF} \oplus \mathbf{c}_3^{EGMF}) \\ \mathbf{p}_l^{MLP} = (\mathbf{l}_0^{MLP} \oplus \mathbf{l}_1^{MLP} \oplus \mathbf{l}_2^{MLP} \oplus \mathbf{l}_3^{MLP}) \\ \mathbf{p}_c^{MLP} = (\mathbf{c}_0^{MLP} \oplus \mathbf{c}_1^{MLP} \oplus \mathbf{c}_2^{MLP} \oplus \mathbf{c}_3^{MLP}) \\ \boldsymbol{\varphi}^{EGMF} = \mathbf{p}_l^{EGMF} \odot \mathbf{q}_c^{EGMF} \\ \boldsymbol{\varphi}^{MLP} = \partial_L(\mathbf{W}_L^T(\partial_{L-1}(\cdots \partial_2(\mathbf{W}_2^T \begin{bmatrix} \mathbf{p}_c^{MLP} \\ \mathbf{q}_c^{MLP} \end{bmatrix} + \mathbf{b}_2) \cdots) + \mathbf{b}_L)) \\ \mathbf{A}_{out} = \text{Softmax}(\begin{bmatrix} \boldsymbol{\varphi}^{EGMF} \\ \boldsymbol{\varphi}^{MLP} \end{bmatrix}) \odot \begin{bmatrix} \boldsymbol{\varphi}^{EGMF} \\ \boldsymbol{\varphi}^{MLP} \end{bmatrix} \\ \hat{y}_{lc} = \sigma(\mathbf{h}^T \mathbf{A}_{out} + \mathbf{b}) \end{cases} \quad (8)$$

式中, \oplus 表示连接, \mathbf{l}_i^{EGMF} 与 \mathbf{l}_i^{MLP} 分别表示 EGMF 和 MLP 模型学习者词向量, $\{\mathbf{l}_0, \mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3\}$ 分别对应 $\{\text{learner_id}, \text{sex}, \text{profession}, \text{job}\}$, \mathbf{c}_i^{EGMF} 与 \mathbf{c}_i^{MLP} 表示 EGMF 和 MLP 模型课程资源词向量。 $\{\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ 分别对应表示 $\{\text{curriculum_id}, \text{name}, \text{complexity}, \text{label}\}$, \mathbf{p}_l^{EGMF} 和 \mathbf{p}_l^{MLP} 分别代表具有辅助信息的学习者特征向量, \mathbf{q}_c^{EGMF} 和 \mathbf{q}_c^{MLP} 分别代表具有辅助信息的课程资源特征向量。

ADCF 模型具体算法描述如下:

Input: 学习者-课程-效用矩阵, 学习者特征矩阵 \mathbf{P} , 课程特征矩阵 \mathbf{Q} , 学习者数量 M , 课程数量 N , 学习者特征潜在因子数量 $F1$, 课程特征潜在因子数量 $F2$, 学习者-课程评价矩阵潜在因子数量 $F3$, Top-k 里面的 k 值, 负采样例个数。

Output: NDCG 和 HR 指标。

Begin

(1) 输入层输入 $\{[\text{learner_id}, \text{sex}, \text{profession}, \text{job}], [\text{curriculum_id}, \text{name}, \text{complexity}, \text{label}]\}$ 的 One-hot 编码作为嵌入层的输入。

(2) 嵌入层分为 EGMF 和 MLP 两部分。EGMF 用于线性特征建模, 嵌入层 EGMF 参数初始化包括激活函数、权重系数、隐含层神经元数等。MLP 用于学习非线性关系特征, 嵌入层 MLP, 参数初始化包括激活函数、隐含层数、隐含层神经元个数等。嵌入层输入的规模从 $(1, 1, n)$ 扩展到了 $(1, m, n)$ 。

(3) 池化层采用 $\max \text{pooling}()$, 输入 $(\mathbf{P_N1} \odot \mathbf{P_N2})$ 和 $(\mathbf{Q_N1} \odot \mathbf{Q_N2})$, 调节他们的规模大小。

(4) 经过 EGMF 和 MLP 特征向量学习之后, 将特征信息进行多模态特征融合, 基于深度学习的多模态特征融合主要是对中间提取到的特征进行拼接或者相加。本文使用的是拼接方式, 即: $\mathbf{X}_n = \text{merge}[(\mathbf{P_N1}$

$\odot P_N2), (Q_N1 \odot Q_N2)]$, 得到输出 X_n 。

(5) 注意力机制层: 多模态融合层的输出 X_n 作为注意力机制的输入, 注意力机制部分本文使用的是 $\text{Softmax}()$ 函数, 通过式(9)获取每个维度特征的权重。

$$\begin{cases} A_n = \text{Softmax}(X_n) \\ A_{out} = A_n \odot X_n \end{cases} \quad (9)$$

式中 X_n 表示 n 维特征向量, A_n 通过 $\text{Softmax}()$ 函数获得注意力计算概率值即关注度, 注意层通过关注度与对应的特征向量相乘获得输出 A_{out} 。

(6) 神经网络层: A_{out} 作为深度神经网络的输入, 损失函数采用成对交叉熵, 各层权重通过反向传播进行调整确定, 评价指标 HR 通过下文的公式(10)计算, NDCG 通过下文的公式(11)计算。最后, 通过一个激活函数(本文选择 Sigmoid 函数)映射至 $[0, 1]$ 范围内, 作为预测结果。

End

3 实验对比

3.1 实验设备与软件环境

实验设备: Huawei Matebook D 笔记本, 处理器为 Intel Core i7 2.2 GHz, 内存为 16 G, 硬盘为 251 G 的 SSD。

软件环境: TensorFlow 1.10, Keras 2.1.5, Pycharm, Numpy 1.13, TensorFlow GPU, CUDA toolkit 8.0, cuDNN v5.1, NVIDIA GTX 1070。

3.2 数据集和评价指标

数据集为网上爬取的经过数据清洗和预处理之后的 2014–2019 年中国大学 MOOC 数据 userlabel08rl, 如表 1 所示。

表 1 userlabel08rl 数据集基本信息

数据集	课程数	人数	交互记录
userlabel08rl	878	3066	207053

采取的评价指标为命中率(hit ratio, HR)和归一化折扣累积增益(NDCG)^[10]。HR 直观地衡量测试项目是否存在于前 10 名列表中, 而 NDCG 通过将较高分指定为顶级排名来计算命中的位置。HR 计算为

$$\text{HitRatio}@k = \frac{\text{NumberOfHits}@k}{|GT|} \times 100\% \quad (10)$$

式中, 分子表示每个用户 Top-k 列表中属于测试集合预测的个数总和, 分母是所有测试集合的个数总和, 分子与分母的比值即为命中率。

NDCG 计算为

$$\begin{cases} CG_k = \sum_{i=1}^k reli \\ DCG_k = \sum_{i=1}^k \frac{2^{reli_i} - 1}{\log_2(i+1)} \\ IDCG_k = \sum_{i=1}^k \frac{1}{\log_2(i+1)} \\ NDCG_k = \frac{DCG_k}{IDCG_k} \end{cases} \quad (11)$$

式中, i 表示处于推荐列表中的位置, $reli$ 表示处于位置 i 的推荐结果的相关性, k 是 Top-k 中的 k 值。

3.3 实验结果

将 ADCF 模型同现有的多种深度协同过滤推荐模型算法, 如 GMF、EGMF、MLP、NeuMF、IUNCF^[11–15] 在 userlabel08rl 数据集上进行实验。为公平起见, 对涉及模型预测因子的算法将模型预测因子统一设为 16。为降低模型评价的时间复杂度, 对评价标准 HR@ k 、NDCG@ k 中的 k , 统一设为 10。

实验结果如图 2~5 和表 2 所示。由图 2~3 可以看出, EGMF 相对于 GMF 在 NDCG@10 和 HR@10 指标上均有提升。图 4~5 展示了不同算法在 userlabel08rl 数据集上迭代次数对模型性能影响的变化情况, 可以看出 ADCF 比 NeuMF 有着较好的性能, 迭代次数到 12 次以后模型就能到达较高的性能, 且性能表现比较稳定。在相同迭代次数下, ADCF 各个指标更好。由此可以认为 ADCF 模型融入注意力机制有利于特征提取, 以及建模能力的提高。

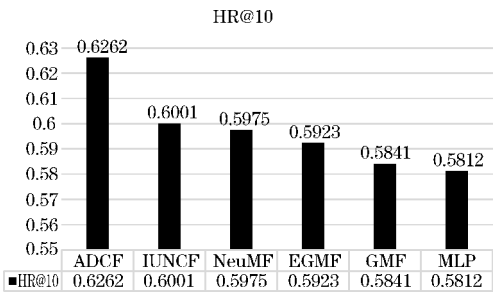


图 2 不同算法的 HR 指标

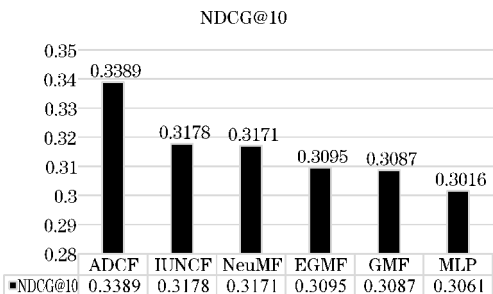


图 3 不同算法的 NDCG 指标

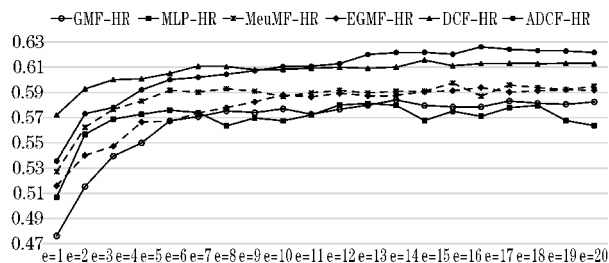


图4 迭代次数对不同算法 HR 指标的影响

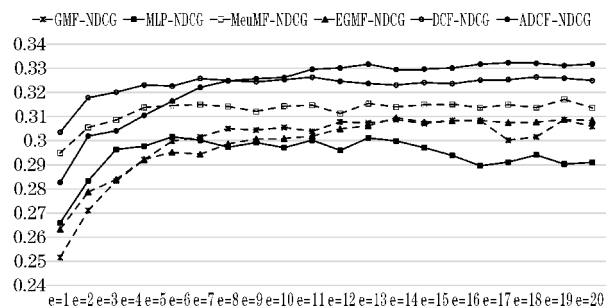


图5 迭代次数对不同算法 NDCG 指标的影响

表2 不同算法在 userlabel08rl 数据集上实验结果

不同算法	HR@10	NDCG@10	t/s
ADCF	0.6262	0.3389	12.9+2.3
IUNCF	0.6001	0.3178	13
NeuMF	0.5975	0.3171	11.2+1.4
EGMF	0.5923	0.3095	8.1
GMF	0.5841	0.3087	7.3
MLP	0.5812	0.3016	11

表2中,ADCF较NeuMF模型在HR@10和NDCG@10两个指标上分别高出2.87%和2.18%,表明考虑辅助信息和集成EGMF的ADCF模型的确能学习到更为关键的高阶特征。ADCF模型同IUNCF模型相比在HR@10及NDCG@10指标上分别高出2.61%与2.05%,表明引入注意力机制和辅助信息的ADCF模型比仅仅考虑辅助信息的IUNCF模型推荐性能更好。由上述实验结果,还可以明显看出集成模型(ADCF、IUNCF、NeuMF)比单个模型(MLP、GMF、EGMF)在两个评价指标上性能表现更优。从时间上分析,ADCF模型框架由于结合了多模态特征融合,引入了注意力思想模型,在训练总的时间上平均每次迭代比IUNCF和NeuMF分别增加了2.2s和2.6s,在训练过程中约需花费12.9s,增加了约1.7s,对测试集完成验证大约需要2.3s,增加了0.7s。尽管引入了注意力思想后的模型比NeuMF模型时间性能下降了一点点,但由于此推荐模型主要用于离线数据进行建模,故影响不大。

4 结束语

提出基于注意力的深度协同过滤(ADCF)在线学

习资源推荐模型,在NeuMF模型基础之上,考虑了辅助信息,集成了拓展广义矩阵,引入注意力机制思想,并将深度学习融入其中,进行建模学习高阶项目之间的关系。在真实的数据集userlabel08rl进行了广泛实验,与现有基于深度学习的推荐框架进行了实验效果对比。实验结果表明,本文的ADCF模型在userlabel08rl真实数据集下的HR@10指标要比其他推荐算法高出2.6%~4.5%,NDCG@10指标要高出个2.1%~3.7%,说明本文所提模型在各个性能指标上均有较大提升。

参考文献:

[1] 孟祥武,纪威宇,张玉杰. 大数据环境下的推荐系统[J]. 北京邮电大学学报, 2015, 38(2): 1-15.

[2] Huang Liwei, Liu Yanbo. Recommendation system based on Deep Learning [J]. Journal of Computer, 2017, 40: 1-28.

[3] ZhangH, Huang T, Lv Z, et al. MCRC: A course recommendation system for MOOCs [J]. Multimedia Tools and Applications, 2017: 1-19.

[4] 黄立威,江碧涛,吕守业,等. 基于深度学习的推荐系统研究综述[J]. 计算机学报,2018, 41(7): 1619-1647.

[5] LIU J, WU C. Deep Learning Based Recommendation: A Survey [M]. Information Science and Applications, 2017:103-109.

[6] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems[C]. Proceedings of the 1st workshop onDeep Learning for Recommender Systems. New York: ACM, 2016: 7-10.

[7] He XN, Liao LZ, Zhang H W, et al. Neural collaborative filtering [C]. Proceedings of the 26th International Conference on World Wide Web. Perth: International World Wide Web Conferences Steering Committee, 2017: 173-182.

[8] Zheng L, Noroozi V, Yu P S. Joint deep modeling of users and items using reviews for recommendation[C]. Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017: 425-434.

[9] Shen Xiaoxuan, Yi Baolin, Zhang Zhaoli, et al. Automatic Recommendation Technology for Learning Resources with Convolutional Neural Network

- [C]. Proceedings of 2016 International Symposium on Educational Technology, Beijing: IEEE, 2016: 30–34.
- [10] Jingyuan Chen, Hanwang Zhang, Xiangnan He, et al. Attentive Collaborative Filtering: Multimedia Recommendation with Item-and Component-Level Attention [C]. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, 2017: 335–344.
- [11] Xiangnan He, Tat-Seng Chua. Neural factorization machines for sparse predictive analytics [C]. Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 2017: 355–364.
- [12] He X, Zhang H, Kan M Y, et al. Fast matrix factorization for online recommendation with implicit feedback [C]. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 2016: 549–558.
- [13] Evangelia Christakopoulou and George Karypis. Local Item-Item Models for Top-N Recommendation [C]. Proceedings of the 10th ACM Conference on Recommender Systems. ACM, 2016: 67–74.
- [14] Bayer I, He X, Kanagal B, et al. A generic coordinate descent framework for learning from implicit feedback [C]. Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017: 1341–1350.
- [15] He X, Chen T, Kan M Y, et al. Trirank: Review-aware explainable recommendation by modeling aspects [C]. Proceedings of the 24th ACM International Conference on Information and Knowledge Management, 2015: 1661–1670.

An Attention-based Deep Collaborative Filtering Model for Online Course Recommendation

FENG Jinhui, TAO Hongcai

(School of Information Science & Technology, Southwest Jiaotong University, Chengdu 611756, China)

Abstract: A recommendation can be thought as a matching problem, i. e. , matching the appropriate item for the appropriate user. Aiming at the problems such as the increasing number of learning platforms and course resources and the poor quality of course recommendation caused by online resource dispersion, this paper integrates the attention mechanism and deep learning into the course recommendation, and proposes an attention-based deep collaborative filtering model for online learning resource recommendation so as to model the high-level course set relation. This model combines the characteristics of learner information and course resources to learn the invisible linear features and nonlinear features of the user and the course, and performs multi-modal feature matching, in order to improve the accuracy of the model in representing the user and the course and the recommendation performance. Through the experiment of crawled learning data from 2014 to 2019 on the MOOC network (MOOC), the results show that the proposed model in this paper is significantly better than other recommendation algorithms in multiple evaluation indexes on the real data set userlabel08rl.

Keywords: deep learning; collaborative filtering; attention mechanism; online courses; recommend model