

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2021.0188

基于学习行为的 MOOC 用户持续学习预测框架

陈辉^{1,*}, 白骏¹, 殷传涛², 荣文戈¹, 熊璋¹

(1. 北京航空航天大学 计算机学院, 北京 100191; 2. 北京航空航天大学 中法工程师学院, 北京 100191)

摘 要: 大型开放式网络课程 (MOOC) 的出现虽然极大地改变了人们的学习方式, 但用户在 MOOC 平台开展学习的学习情况及完成率预测仍是目前一个重要的技术挑战。针对预测的需求, 从用户的学习行为中对用户和课程进行分析, 采用长短时记忆机对学习者的学习活动进行建模, 采用多头注意力机制对用户和课程之间的交互活动情况进行分析, 提出一个基于门控单元的特征融合框架, 用于学习情况预测。在公开数据集上的结果表明: 所提框架能够提升预测精度, 使得 MOOC 平台能够尽可能早地对用户活动进行干预, 从而提升整体的 MOOC 平台使用体验。

关 键 词: 大型开放式网络课程; 预测框架; 用户; 课程内容; 学习行为

中图分类号: TP399; G434

文献标志码: A

文章编号: 1005-5965(2023)01-0074-09

大型开放式网络课程 (massive open online courses, MOOC) 作为目前一种最重要的在线学习机制之一, 极大地改变了学习者的学习行为, 使得不同的学习者能够不受时间和空间的限制, 随意进行感兴趣知识的学习, 因此, MOOC 在当今网络学习大潮中扮演越来越重要的角色, 受到越来越多的欢迎^[1]。然而, 作为一种新的学习机制, MOOC 的发展遇到了一系列的挑战, 其中一个最重要的问题是, 有相当多的学习者无法完成 MOOC 平台上的课程学习, Lykourantzou 等^[2]通过研究显示, 只有不到 10% 的注册用户可以在 MOOC 平台上完成特定的课程学习。

对于任何一个新的信息化系统, 其成功的要素一般可以从用户使用过程中的愉悦度、参与度、接受度、留存率及任务完成度等角度来研究, 该 HEART 度量框架是一个广泛适用的评价体系^[3]。因此对于 MOOC 平台, 如何吸引学习者能够留在 MOOC 平台中不断学习, 去完成一些特定课程, 提高留存率和任务完成度, 是一个非常重要的任务, 尽管简

单地使用退课率来评估 MOOC 平台对于学习者的学习效果可能并不一定准确, 但其确实反映了一些具有挑战性的问题且值得进一步研究。

学习者的学习习惯对于学习的效果会有直接的关联, 不同的课程内容对于学习的效果也有不同的作用。因此, 更好地了解学习者对于不同课程的学习特征对于 MOOC 平台的退课预测研究有着重要的价值。对于学习者来说, 有很多因素可能会影响注册用户在 MOOC 平台上学习课程的持久性, 如年龄、性别、以往的学习经历、受教育程度等^[2], 此外, 学习者在 MOOC 平台上的一些学习习惯, 也可能对其在平台上的学习效果产生重要的影响, 比如在学习过程中尝试提交作业或其他测试的行为^[4], 对论坛帖子的浏览和关注^[5], 课程视频的点击^[6]等。同样的, 不同类型的课程, 如课程内容上的差异^[7]等, 也可能会有不同的退课表现。因此, 本文在研究 MOOC 平台退课率时, 不仅要研究学习者学习习惯带来的影响, 也要研究课程本身带来的影响。基于这样的思考, 本文提出了一个基于用户和课程

收稿日期: 2021-04-12; 录用日期: 2021-04-21; 网络出版时间: 2021-04-23 16:29

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210423.1555.002.html

基金项目: 国家自然科学基金 (61977003)

*通信作者: E-mail: chenhui@buaa.edu.cn

引用格式: 陈辉, 白骏, 殷传涛, 等. 基于学习行为的 MOOC 用户持续学习预测框架 [J]. 北京航空航天大学学报, 2023, 49 (1): 74-82.

CHEN H, BAI J, YIN C T, et al. Behavior based MOOC user dropout predication framework [J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49 (1): 74-82 (in Chinese).

感知增强的活动学习框架。在该框架中,首先提出了一种语义感知编码机制来分别对用户和课程的活动进行编码,然后使用注意机制分别从用户和课程的角度提取学习活动的相关性,最后设计了一个基于门控机制的上下文提取机制,将用户和课程的特征整合在一起,以实现最终的退课行为预测。为了评价所提框架,本文在 XuetangX 数据集上^[8]进行了实验研究,实验结果验证了本文提出的退课预测框架的可行性。

1 MOOC 退课率预测现状

近年来,MOOC 平台为全世界各地的数百万用户提供了众多的免费在线课程^[9],一些 MOOC 平台,例如 Coursera、udacity、edx、XuetangX 等已经成为开放式和远程教育最重要的在线学习平台^[1]。尽管 MOOC 的优势已经得到充分证明,但值得注意的是较高的退课率已成为 MOOC 实际应用中最重要的一個挑战^[10],有研究表明,超过 90% 的用户都不能顺利地在 MOOC 平台上完成课程^[11]。因此,预测学习者放弃课程的可能性就变得非常重要^[12],从而使得 MOOC 平台能够及时提供支持性干预措施,通过各种不同的激励方法来鼓励学习者继续完成课程的学习^[13]。

通过分析用户在 MOOC 平台上学习课程方面的特点,预测其潜在的退课可能性是在线教育领域一个非常重要的课题^[14],2015 年 KDD 大会还特地为此类退课率预测挑战举办了数据挖掘的竞赛。为了降低退课率,需要给学习者提供一个满意的 MOOC 平台学习体验,而好的学习体验与很多因素相关^[15],例如:学习者本身背景、课程内容、平台的质量、学校的特点等。有学习者通过研究指出,MOOC 平台上学习者退课率高的主要因素和学习者的学习行为与 MOOC 平台属性相关,其中与学习者相关的属性包括学习的动机、学习时间、背景知识和技能等因素,而与 MOOC 平台相关的属性包括课程设计和课程互动性等因素^[7]。类似的,Goopio 和 Cheung^[16]的研究表明,MOOC 平台上的退课率主要与学习体验、互动性、课程设计、技术、语言、时间和情境七大因素有关。

为了降低 MOOC 平台学习者的退课率,一个可行的解决办法是尽早预测退课行为,从而使得平台能够提前部署一些干预手段,帮助和鼓励学习者来完成课程^[17]。不同学习者的不同学习行为、学习习惯和学习时间都会影响学习者的退课行为,这其中学习者的学习行为对于退课的预测有着非常重要的作用^[14]。在 MOOC 平台上,通过研究分析学习者

的学习行为来预测退课的可能性一直是退课率预测领域的重要挑战。例如,文献^[18-19]对学生参与 MOOC 课程的情况进行了调查,开发了一种潜在动态因子图(LadFG)来建模和预测学习行为,同时,进一步提出融合特征提取、特征选择和学习者分类的预测模型,并利用 MOOC 平台提供的一些学习行为特征进行预测。

在研究 MOOC 平台的学习者学习行为特征的工作中,最常用的一种学习行为特征是注册用户在 MOOC 中学习模式的点击流。点击流是网络课程研究中一个重要的学习者行为的时间序列指标^[20],通过使用点击流数据可以构建一个不断变化的学习者学习状态表示,利用这个学习状态表示,就能对学习者的行为进行跟踪,从而预测其可能出现的退课行为。点击流是一种重要的行为模式,并且是评估 MOOC 学习成果的一种重要数据来源^[21-22],因为不同的学生通常有不同的学习方式,这意味着对于同一门课程,不同类型的学习者可能会有不同的活动和学习结果^[23]。为了更好地了解点击流,需要更好地对点击流进行建模,较早的尝试提取用户操作方面的特征,如不同点击类型的计数^[24],虽然这种方法显示出一定的效果并且易于实现,但无法研究点击流所包含的一些复杂属性,比如点击之间的顺序^[25]。目前随着深度学习技术的发展,出现了许多基于神经网络的方法来对点击流进行建模。

为了充分利用点击流中的顺序信息,一些研究人员建议使用序列感知学习技术来研究学习者学习模式,进而预测学习者的退课行为。例如,Wang 等^[26]提出一种基于递归神经网络(recurrent neural network, RNN)的深度模型用于退课预测。类似的,Tang 等^[27]提出一种基于 RNN 具有短期记忆单元的方法。Xiong 等^[28]则进一步开发了基于长短期记忆(long short-term memory, LSTM)网络的方法来预测学习状态。同样,Fei 和 Yeung^[25]提出一个 LSTM 单元的神经网络模型进行退课预测^[29]。这种针对学习者学习行为序列的建模方法在预测退课行为时显示出了巨大的潜力,因为点击流确实具有顺序属性。但是,点击流活动序列内不同活动之间的内在联系也值得充分研究。例如,与其他活动相比,“加载视频”与“观看视频”的相关性更高。因此,有学者指出,除了点击流内部活动之间的时序特性,部分点击流内部活动之间的依赖性应该也有助于提高总体退学预测任务的准确性。

在研究点击流内部活动关系方面,Wang 等^[26]提出了一种结合了卷积神经网络(convolutional neural networks, CNN)和 RNN 方法的 ConRec 网络,从原始 MOOC 数据中自动提取特征。类似的,为了利

用学习行为之间的相关性, Wen 等^[23]设计了一个基于时间序列的二维矩阵作为 CNN 模型的输入, 将时间信息与学习者的行为特征相结合, 利用时间序列解决行为特征的局部相关性问题。除了研究活动之间的相关性以外, 还需要指出不同用户和不同课程可能具有不同的活动模式, 这意味着活动相关性也应分别从用户和课程的角度进行研究。实际上, 一些研究人员指出, 从用户和课程不同的角度去研究活动相关性也将影响 MOOC 课程的退课率预测^[30], 因为不同的课程可能与活动具有不同的相关模式, 在学习行为的研究中, 不同种类的课程对学生的行为可能有不同的影响^[15]。Feng 等^[8]针对这个问题进行了研究, 提出了一种上下文感知功能交互网络(context-aware feature interaction network, CFIN)来预测辍学概率, 并在更大的数据集上取得了最好的结果。

2 退课预测框架

2.1 问题定义

令 $u \in U$ 和 $c \in C$ 为用户和课程, 其中 U 和 C 分别为用户集合和课程集合, 从用户 u 学习周期为 D_u 的 c 课程学习日志中提取用户的学习行为 $X(u, c)$, 其中每个元素 $x_i(u, c)$ 表示特定的行为, v_i 是 $x_i(u, c)$ 的归一化统计值, 其相应的行为名称为 $s_i = \{w_{ij}\}_{j=1}^{l_{s_i}}$ (如“加载视频”), 其中 w_{ij} 表示行为 $x_i(u, c)$ 名称中的第 j 个

单词, l_{s_i} 为该名称的长度。学习行为的上下文信息包含用户信息 $Z(u)$ 和课程信息 $Z(c)$, 其中用户信息元 $z_i(u)$ 包含学习者的个人基本信息(如年龄、性别、受教育程度、注册课程数量等), 课程信息元素 $z_i(c)$ 则包含课程基本信息(如课程种类及注册课程的用户数)。

退课预测任务可根据上述变量形式化定义为: 给定过去时间段内发生的学习行为 $X(u, c)$ 及相关上下文信息 $Z(u)$ 和 $Z(c)$, 预测用户 u 是否会在预测周期 D_p 内放弃课程 c 的学习, 用于是否退课的结果是一个二分类的值 $y_{(u, c)}$, 用 1 表示用户在预测周期内没有放弃学习任务, 否则为 0, 因此本文目的是学习一个预测函数, 使得其预测结果 $\hat{y}_{(u, c)}$ 尽可能地逼近真实结果 $y_{(u, c)}$ 。

2.2 框架结构

针对 MOOC 平台学习者的学习行为, 本文首先提取学习行为的语义信息, 并捕捉学习行为之间的交互关系。在此基础上, 进一步从用户和课程的角度分别对行为进行分析, 获得从用户和课程角度观察到的行为信息, 最后, 通过一个门控机制将用户、课程信息及与用户、课程相关的行为交互特征进行融合, 并最终用于用户退课的预测, 因此, 本文提出一个基于用户和课程信息感知增强的行为学习框架, 如图 1 所示。

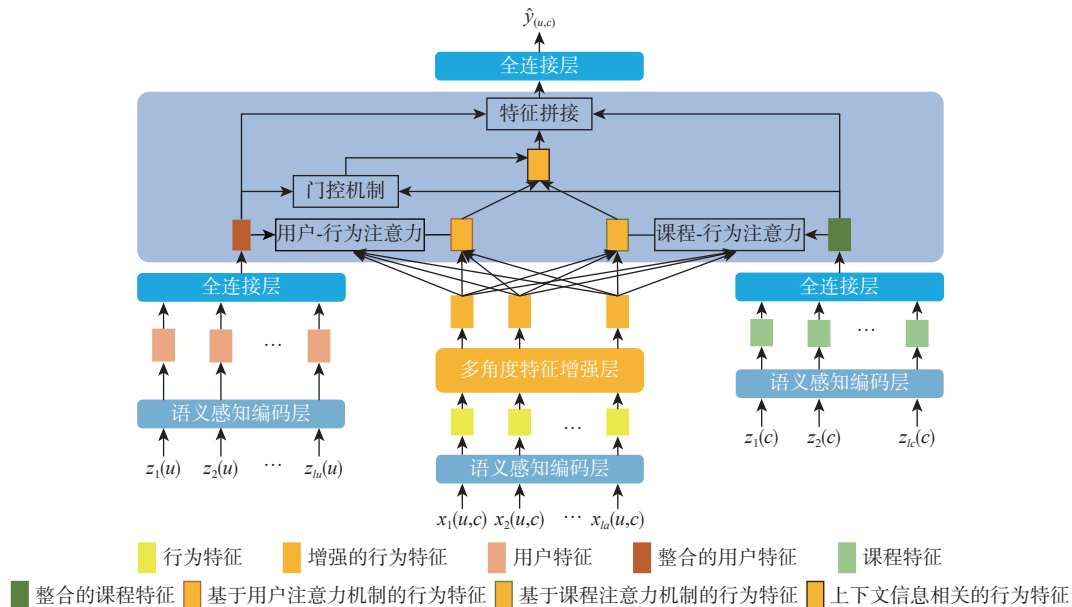


图 1 基于用户和课程信息感知增强的退课预测框架

Fig. 1 User and course information perception enhancement based dropout prediction framework

2.2.1 用户和课程的语义编码

为了构建一个有效的退课预测框架, 本文需要设计一个有效的方法来表示上下文信息和学习行为。早期工作一般直接使用随机初始化的嵌入层

或学习行为统计值作为表示层, 而忽略了此类信息的语义。因此, 本文提出语义感知编码层(semantic aware encoding, SAE), 从上下文信息和学习行为中提取语义感知特征。

本文以学习行为的编码为例,如图 2 所示。给定名称为 $s_i = \{w_{ij}\}_{j=1}^{l_{s_i}}$ 及归一化统计值为 v_i 的学习行为,使用 GloVe 词嵌入将 s_i 中的每个词映射到向量空间并对统计值进行归一化。使用 LSTM 作为编码器,从 s_i 的词嵌入中提取行为对应的语义。LSTM 在 t 时刻的输入门 z_t^I 、遗忘门 z_t^F 、输出门 z_t^O 分别计算如下:

$$z_t^I = \sigma(W_I[g(w_{it}); \mathbf{h}_{t-1}] + \mathbf{b}_I) \quad (1)$$

$$z_t^F = \sigma(W_F[g(w_{it}); \mathbf{h}_{t-1}] + \mathbf{b}_F) \quad (2)$$

$$z_t^O = \sigma(W_O[g(w_{it}); \mathbf{h}_{t-1}] + \mathbf{b}_O) \quad (3)$$

式中: W_I 、 W_F 、 W_O 和 $\mathbf{b}_I$ 、 $\mathbf{b}_F$ 、 $\mathbf{b}_O$ 为 LSTM 中可学习的参数; $g(w_{it})$ 为词 w_{it} 对应的 GloVe 词嵌入; $\mathbf{h}_{t-1}$ 为 $t-1$ 时刻的隐状态; $[a; b]$ 为张量 a 和 b 的拼接结果; σ 为 sigmoid 激活函数。候选单元状态 $\tilde{c}_t$ 为

$$\tilde{c}_t = \tanh(W_C[g(w_{it}); \mathbf{h}_{t-1}] + \mathbf{b}_C) \quad (4)$$

式中: W_C 、 $\mathbf{b}_C$ 分别为用于计算候选单元状态线性层的映射矩阵和偏置; 下标 C 为单元状态。

使用输入门和遗忘门计算单元状态,其中 $\odot$ 表示张量的按位乘法运算:

$$\mathbf{c}_t = z_t^I \odot \tilde{c}_t + z_t^F \odot \mathbf{c}_{t-1} \quad (5)$$

t 时刻的隐状态 $\mathbf{h}_t$ 为

$$\mathbf{h}_t = z_t^O \odot \tanh(\mathbf{c}_t) \quad (6)$$

最后时刻的隐状态将作为学习行为的语义特征,将其乘以归一化后的统计值得到最终的学习行为特征表示 $\mathbf{f}_i^a$ 为

$$\mathbf{f}_i^a = v_i \odot \mathbf{h}_{l_{s_i}} \quad (7)$$

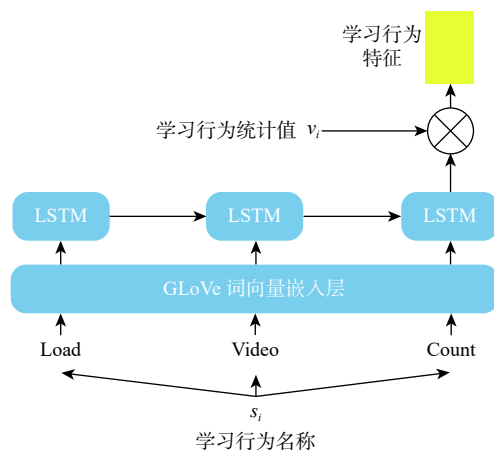


图 2 基于语义编码的点击活动特征编码

Fig. 2 Semantic oriented based click activity feature encoding

2.2.2 多角度特征增强

在退课预测任务中,有一些学习行为是相互关联的,如“加载视频”和“播放视频”有时序先后关系。这类相关性可以帮助框架更深入地了解用户

行为。为了使框架具备对学习行为之间相互作用的建模能力,本文提出了多角度特征增强层(multi-perspective feature enhancement, MFE),该层从多个角度提取了学习行为之间的相关性。

令 $\mathbf{F}^a \in \mathbf{R}^{l_a \times d_c}$ 表示学习行为特征张量, $\mathbf{f}_i^a$ 表示 $\mathbf{F}^a$ 中第 i 类行为所对应的分量,其中 l_a 和 d_c 分别表示学习行为的数量和学习行为特征张量的长度。为了从多角度提取学习行为之间的相关性,本文采用了多头缩放点积自注意力机制,每个头的特征 $\mathbf{H}_i$ 用于提取单个角度的相关性,其计算式如下:

$$\mathbf{H}_i = \text{softmax} \left(\frac{\mathbf{F}^a (\mathbf{F}^a)^T}{\sqrt{d_h}} \right) \mathbf{F}^a \quad (8)$$

式中: d_h 为 $\mathbf{H}_i$ 特征对应的维度。随后将所有的 $\mathbf{H}_i$ 特征进行拼接得到多角度行为相关性特征 $\tilde{\mathbf{H}}$:

$$\tilde{\mathbf{H}} = [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_{l_h}] \quad (9)$$

式中: l_h 为注意力多头数,本文利用线性层整合不同的头特征,并分别使用层归一化和残差连接来稳定特征比例和梯度。于是获得了具有多角度相关性增强的学习行为特征 $\tilde{\mathbf{F}}^a$:

$$\tilde{\mathbf{F}}^a = \text{LayerNorm}(\mathbf{F}^a) \quad (10)$$

2.2.3 基于门控机制的上下文相关特征优化

每个学习行为的重要性会根据用户和课程的不同而变化,因此,直接使用所有学习行为的特征会存在很多噪声从而导致预测性能较差。为了融合学习行为的相关上下文信息来优化行为特征,本文提出了基于门控机制的上下文相关特征提取优化层(gated context-related feature refine, GCR),该层分别基于用户信息特征和课程信息特征得到带有相关注意力机制的行为特征,然后使用门控机制合并这 2 个特征。

令 $\mathbf{F}^u \in \mathbf{R}^{l_u \times d_c}$ 和 $\mathbf{F}^c \in \mathbf{R}^{l_c \times d_c}$ 分别表示从语义感知编码层得到的用户信息特征和课程信息特征,其中 l_u 和 l_c 分别为用户和课程包含的特征种类数量,本文利用全连接层将 $\mathbf{F}^u$ 和 $\mathbf{F}^c$ 分别整合成单个特征张量:

$$\tilde{\mathbf{f}}^u = \mathbf{W}_{P_u} \text{Flatten}(\mathbf{F}^u) + \mathbf{b}_{P_u} \quad (11)$$

$$\tilde{\mathbf{f}}^c = \mathbf{W}_{P_c} \text{Flatten}(\mathbf{F}^c) + \mathbf{b}_{P_c} \quad (12)$$

式中: $\mathbf{W}_{P_u}$ 和 $\mathbf{b}_{P_u}$ 分别为用于整合用户信息特征的线性层的映射矩阵和偏置; $\mathbf{W}_{P_c}$ 和 $\mathbf{b}_{P_c}$ 分别为用于整合课程信息特征的线性层的映射和偏置; Flatten 表示将多维特征铺平成一维特征,从而能够将其输入全连接层进行运算。提取基于用户注意力机制的行为特征,上述被整合的用户特征 $\tilde{\mathbf{f}}^u$ 与每个增强的学习行为特征 $\tilde{\mathbf{f}}_i^a$ 进行拼接,用于计算用户-行为注意力得分 α_i :

$$\hat{\alpha}_i = \mathbf{W}_{\text{Att2}} \text{RELU}(\mathbf{W}_{\text{Att1}} [\tilde{\mathbf{f}}^u; \hat{\mathbf{f}}_i^a]) \quad (13)$$

$$\alpha_i = \frac{\exp(\hat{\alpha}_i)}{\sum_{j=1}^{l_a} \exp(\hat{\alpha}_j)} \quad (14)$$

式中: $\mathbf{W}_{\text{Att2}}$ 、 $\mathbf{W}_{\text{Att1}}$ 分别为用于计算用户-行为注意力得分的第1层、第2层线性层的映射矩阵。

用户-行为注意力得分 α_i 体现了学习行为 $x_i(u, c)$ 对于用户的重要性程度,进一步使用注意力得分对增强的学习行为特征进行加权求和,从而得到基于用户注意力机制的行为特征 $\hat{\mathbf{f}}_u^a$:

$$\hat{\mathbf{f}}_u^a = \sum_{i=1}^{l_a} \alpha_i \hat{\mathbf{f}}_i^a \quad (15)$$

可以采用相同的方式得到基于课程注意力机制的行为特征 $\hat{\mathbf{f}}_c^a$,现在本文有了 $\hat{\mathbf{f}}_u^a$ 和 $\hat{\mathbf{f}}_c^a$ 这2个特征,由于用户信息和课程信息对于不同实例的重要性不同,因此进一步采用门控机制来合并这些特征。利用整合的用户特征 $\tilde{\mathbf{f}}^u$ 和课程特征 $\tilde{\mathbf{f}}^c$ 计算门控得分的实数值,将其用于权衡基于用户注意力机制的行为特征和基于课程注意力机制的行为特征:

$$g = \sigma(\mathbf{W}_G [\tilde{\mathbf{f}}^u; \tilde{\mathbf{f}}^c]) \quad (16)$$

式中: $\mathbf{W}_G$ 为用于计算门孔得分的线性层映射矩阵。

根据门控得分融合 $\hat{\mathbf{f}}_u^a$ 和 $\hat{\mathbf{f}}_c^a$ 得到与上下文信息相关的行为特征:

$$\tilde{\mathbf{f}}^a = g \odot \hat{\mathbf{f}}_u^a + (1 - g) \odot \hat{\mathbf{f}}_c^a \quad (17)$$

将整合的用户特征 $\tilde{\mathbf{f}}^u$,课程特征 $\tilde{\mathbf{f}}^c$ 和上下文信息相关的学习行为特征 $\tilde{\mathbf{f}}^a$ 进行拼接作为分类特征,传入线性层和sigmoid激活函数获得预测的退学概率:

$$\hat{y}_{(u,c)} = \sigma(\mathbf{W}_P [\tilde{\mathbf{f}}^u; \tilde{\mathbf{f}}^c; \tilde{\mathbf{f}}^a] + \mathbf{b}_P) \quad (18)$$

式中: $\mathbf{W}_P$ 和 $\mathbf{b}_P$ 分别为用于计算退学概率的线性层映射矩阵和偏置。

在训练阶段,本文使用交叉熵损失函数来学习所有的框架参数:

$$\text{Loss} = -y_{(u,c)} \ln(\hat{y}_{(u,c)}) - (1 - y_{(u,c)}) \ln(1 - \hat{y}_{(u,c)}) \quad (19)$$

3 实验验证

3.1 数据集

为了验证所提出的框架,在本文研究中,对XuetangX数据集的教师定型模式课程进行了实验。XuetangX数据集发布于2019年,包含超过220 000个实例,并且比KDDCUP 2015大得多,KDDCUP 2015也来自XuetangX数据集,且包含超过120 000个活

动日志。本文通过预处理并将整个数据集拆分为142 149个训练实例,15 794个开发实例和67 699个测试实例,与文献[8]采用了相同的配置,历史期和预测期分别为35 d和10 d。

3.2 实验设置

本文根据开发集上框架的性能选择超参数。选择与文献[8]相似的配置,使用50维的GloVe嵌入向量和1个包含50个隐藏单元的LSTM层。此外,使用了1层4头自注意力层,使用Adam优化器对框架进行了优化,并将学习率设置为 10^{-4} 。权重为 10^{-5} 的L2正则化还可以避免过度拟合。将批次大小设置为32,训练了50个epoch,使用在开发集上表现最佳的框架参数来进行评估。

3.3 对比方法

为了验证所提框架,本文将其与其他方法进行比较。

1) 逻辑回归(logistics regression, LR): LR是一种广泛用于分类的线性回归分析方法。能够基于对不同变量之间的回归系数的分析来估计某个事件是否会发生。LR已作为基本技术^[31-32]应用于许多MOOC退课预测任务中。

2) 支持向量机(support vector machine, SVM): SVM广泛用于二分类任务,并且能够利用从点击流^[33]中精心选择的特征来预测MOOC退课。

3) 随机森林(random forest, RF): RF是基于决策树的方法,在训练过程中引入随机性来提高分类器的性能,也是一种MOOC退课预测^[34]中广泛使用的基本方法。

4) 梯度提升决策树(gradient boosting decision tree, GBDT): 基于GBDT的退课预测模型是使用决策树的整体来预测标签,深度大于1,并且在此任务中显示了可喜的结果^[35]。

5) 全连接神经网络(deep neural network, DNN): 1个3层DNN,用于研究将深度学习技术用于退课预测任务的可行性。

6) CFIN: 上下文感知功能交互网络^[8],是目前面向较大的XuetangX数据集的重要模型。

3.4 评估指标

本文使用精度(ACC)^[36]、接受者操作特征曲线(ROC)、曲线下的面积(AUC)^[37]来评估所提框架的性能。对于二分类问题,根据预测分类和实际标签可以将样本划分为4类,即真正例(TP)、假正例(FP)、真反例(TN)和假反例(FN)。分类结果可以用“混淆矩阵”表示,如表1所示。

分类框架中每个样本的输出结果一般为实数值,对于设定的阈值,将该实数值和阈值比较,大于

表 1 分类结果混淆矩阵

Table 1 Confusion table confusion matrix for classification results

实际标签	预测分类	
	正例数量	反例数量
正例数量	TP	FN
反例数量	FP	TN

阈值的为正样本,反之为负样本。通过设定不同的阈值,可以得到不同的分类结果。将样本按照预测得到的实数值从高到低进行排序,将每个样本对应的实数值作为阈值,就可以得到和样本数量相当的分类结果。考虑不同阈值下的分类结果,综合反映了分类框架得到结果的排序质量,从而能够更好地反映框架的泛化能力。根据实际标签,可以计算出假阳率(FPR)及真阳率(TPR)的值:

$$FPR = \frac{FP}{FP + TN} \quad (20)$$

$$TPR = \frac{TP}{TP + FN} \quad (21)$$

根据 FPR 和 TPR, ACC 可定义为

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (22)$$

ROC 也可以利用 FPR 和 TPR 来计算,将每个阈值对应的 (FPR, TPR) 点绘制在坐标图中,得到 ROC 曲线,ROC 曲线的横坐标为假正例率,纵坐标为真正例率。利用 ROC 曲线对分类框架效果进行比较时,若各框架的 ROC 曲线没有发生交叉,直观上来讲,最外层的 ROC 曲线包覆住了其他框架的 ROC 曲线,因此,最外层的 ROC 曲线对应的分类框架效果最好。若框架的 ROC 曲线发生了交叉,那么通常采用比较 ROC 曲线下面积的方式^[37]。因此,该面积即 AUC 被采纳为通用的分类框架优劣判据。在实际训练过程中一般只有有限个样本,因此绘制得到的 ROC 曲线由有限个 (FPR, TPR) 点连接而成,点坐标集合为: $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, 则 AUC 可以采用将黎曼积分化为求和形式的方法求解:

$$AUC = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i)(y_i + y_{i+1}) \quad (23)$$

3.5 结果与分析

表 2 列出了有关 ACC 和 AUC 的总体性能,就 AUC 而言,本文提出的框架可以胜过大多数机器学习方法,但要比最先进的方法 CFIN 要低一些。而在 ACC 方面,本文提出的框架可以胜过 CFIN。表 2 表明,本文提出的框架与大型数据集中的复杂方法相比具有竞争优势。

为了进一步分析提出的 SAE、MFE 和 GCR 组件

表 2 ACC 和 AUC 总体性能

Table 2 Overall performance in terms of ACC and AUC

方法	ACC/%	AUC/%
LR		82.23
SVM		82.86
RF		83.11
DNN		85.64
GBDT		85.18
CFIN	84.78	86.40
本文方法	85.09	86.07

的效果,本文还通过以下方法进行消融研究:①使用随机初始化的特征向量而不是语义感知编码层(不包含 SAE);②删除多视角特征增强层(不包含 MFE);③用平均池化层替换增强的学习行为特征(不包含 GCR)上门控的上下文相关特征优化层。本文注意到最明显的下降发生在没有 MFE 的情况下,这证明了捕获学习行为之间交互关系的重要性。在没有 SAE 的情况下,所获得的框架特征语义要比具有随机初始化特征的框架更好。如表 3 所示,门控上下文相关特征细化层可以帮助框架过滤不相关的活动特征,并根据不带 GCR 的结果提高性能。

表 3 消融实验

Table 3 Ablation test

方法	ACC/%	AUC/%
本文方法	85.09	86.07
w/o SAE	84.97	85.96
w/o MFE	84.91	85.99
w/o GCR	84.91	86.00

注: w/o 表示消融实验中的 without。

3.6 案例与分析

本文对一个测试样本进行采样,并可视化其学习行为-学习行为注意力得分、上下文信息-学习行为注意力得分。

如图 3 所示,学习行为与学习行为之间的注意力系数热度图,可以看到框架可以捕捉到一些具有明显时序关系的学习行为之间的交互关系,如“点击论坛”与“问题获取”、“问题获取”与“创建评论”之间的注意力分数明显要更高。而对一些相关性较低的学习行为组合则分配了较低的注意力分数,如“播放视频”和“关闭课件”。

如图 4 所示,上下文信息与学习行为之间的注意力系数热度图,可以看到用户功能主要集中在学习行为“问题获取”和“创建评论”上,而课程功能则关注于学习行为“点击课件”、“关闭课件”及“载入视频”等,这有助于框架过滤噪声和提取与上下文最相关的学习行为特征。

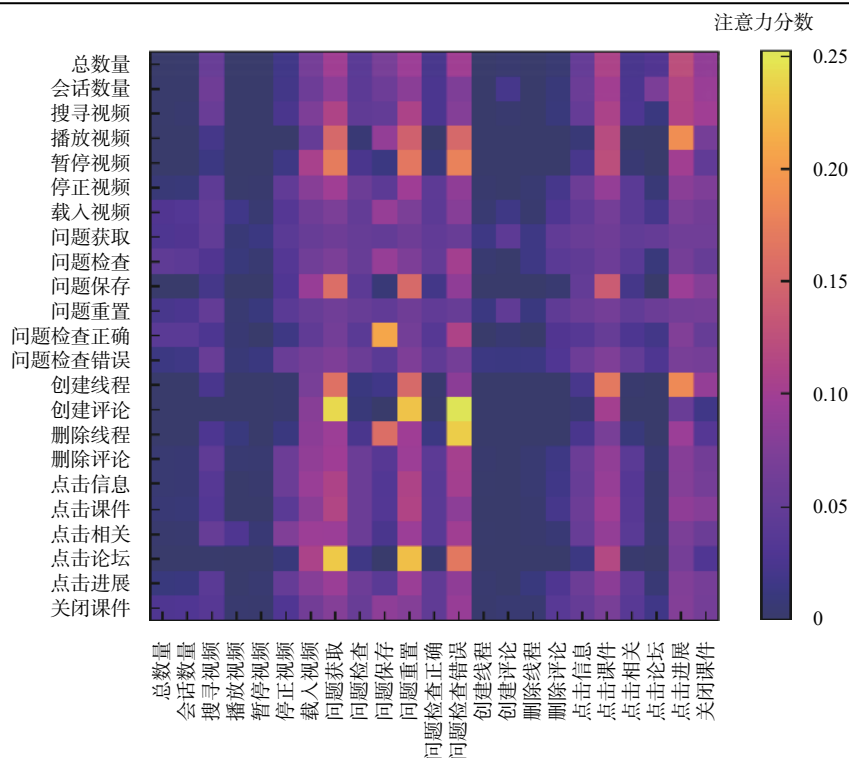


图3 学习行为-学习行为注意力热度图

Fig. 3 Study activity-study activity attention heat map

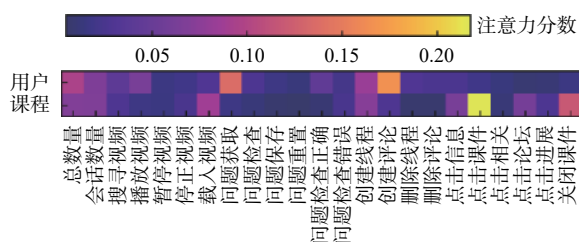


图4 上下文信息-学习行为注意力热度图

Fig. 4 Context-study activity attention heat map

4 结 论

针对 MOOC 平台较高退课率的问题, 本文提出了一个基于长短时记忆机和多头注意力机制的 MOOC 平台用户学习情况预测框架, 在该框架中:

1) 提出语义编码机制, 将用户的学习行为进行分析, 为用户和相应课程进行编码, 可以更好地获得用户和课程的特征。

2) 提出一个多角度的用户和课程之间交互的编码方法, 更好地将用户和课程之间的交互关系进行分析。

3) 提出基于门控机制的上下文相关行为特征优化机制, 最后将获得的用户、课程及优化后的行为交互特征融合并用于退课预测, 并在 XuetangX 数据集上进行了实验验证。

4) 所提框架能获得较好的总体性能, 消融实验也验证了本文提出的不同编码机制和特征融合机制的有效性。本文所提方法能够使得 MOOC 平台

能够尽可能早的对用户活动进行干预, 从而提升整体的 MOOC 平台使用体验。

参考文献 (References)

- [1] MARTIN F G. Will massive open online courses change how we teach?[J]. *Communications of the ACM*, 2012, 55(8): 26-28.
- [2] LYKOURANTZOU I, GIANNOUKOS I, NIKOLOPOULOS V, et al. Dropout prediction in E-learning courses through the combination of machine learning techniques[J]. *Computers & Education*, 2009, 53(3): 950-965.
- [3] FISHBEIN M, AJZEN I. Belief, attitude, intention, and behavior: An introduction to theory and research[J]. *Journal of Business Venturing*, 1977, 5: 177-189.
- [4] CHEN C, SONNERT G, SADLER P M, et al. Computational thinking and assignment resubmission predict persistence in a computer science MOOC[J]. *Journal of Computer Assisted Learning*, 2020, 36(5): 581-594.
- [5] CROSSLEY S A, PAQUETTE L, DASCALU M, et al. Combining click-stream data with NLP tools to better understand MOOC completion[C]//*Proceedings of the 6th International Conference on Learning Analytics & Knowledge*. New York: Association for Computing Machinery, 2016: 6-14.
- [6] GARDNER J, BROOKS C. Dropout model evaluation in MOOCs[C]//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2018: 7906-7912.
- [7] DALIPI F, IMRAN A S, KASTRATI Z. MOOC dropout prediction using machine learning techniques: Review and research challenges[C]//*Proceedings of 2018 IEEE Global Engineering Education Conference*. Piscataway: IEEE Press, 2018: 1007-1014.
- [8] FENG W, TANG J, LIU T X. Understanding DROPOUTs in MOOCs[C]//*Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2019: 7906-7912.

- cial Intelligence. Palo Alto: AAAI Press, 2019: 517-524.
- [9] ALRAIMI K M, ZO H, CIGANEK A P. Understanding the MOOCs continuance: The role of openness and reputation[J]. *Computers & Education*, 2015, 80: 28-38.
- [10] ITANI A, BRISSON L, GARLATTI S. Understanding learner's drop-out in MOOCs[C]//*Proceedings of 19th International Conference on Intelligent Data Engineering and Automated Learning*. Berlin: Springer, 2018: 233-244.
- [11] YANG D, WEN M, HOWLEY I K, et al. Exploring the effect of confusion in discussion forums of massive open online courses[C]//*Proceedings of the 2nd ACM Conference on Learning @ Scale*. New York: Association for Computing Machinery, 2015: 121-130.
- [12] YANG D, SINHA T, ADAMSON D, et al. "Turn on, tune in, drop out": Anticipating student dropouts in massive open online courses[C]//*Proceedings of 2013 NIPS Workshop on Data Driven Education*. New York: Curran Associates, 2013.
- [13] JEON B, PARK N. Dropout prediction over weeks in MOOCs by learning representations of clicks and videos[C]//*Proceedings of 2020 AAAI Workshop on Artificial Intelligence for Education*. Palo Alto: AAAI Press, 2020.
- [14] JIN C. MOOC student dropout prediction model based on learning behavior features and parameter optimization[J/OL]. *Interactive Learning Environments*, 2020 (2020-08-12)[2021-02-12]. <https://doi.org/10.1080/10494820.2020.1802300>.
- [15] ADAMOPOULOS P. What makes a great MOOC? An interdisciplinary analysis of student retention in online courses[C]//*Proceedings of the 2013 International Conference of Information Systems*. Milano: AIS eLibrary, 2013.
- [16] GOOPIO J, CHEUNG C. The MOOC dropout phenomenon and retention strategies[J]. *Journal of Teaching in Travel & Tourism*, 2021, 21(2): 177-197.
- [17] XING W, DU D. Dropout prediction in MOOCs: Using deep learning for personalized intervention[J]. *Journal of Educational Computing Research*, 2019, 57(3): 547-570.
- [18] QIU J, TANG J, LIU T X, et al. Modeling and predicting learning behavior in MOOCs[C]//*Proceedings of the 9th ACM International Conference on Web Search and Data Mining*. New York: Association for Computing Machinery, 2016: 93-102.
- [19] QIU L, LIU Y, HU Q, et al. Student dropout prediction in massive open online courses by convolutional neural networks[J]. *Soft Computing*, 2019, 23(20): 10287-10301.
- [20] JEON B, SHAFRAN E, BREITFELLER L, et al. Time-series insights into the process of passing or failing online university courses using neural-induced interpretable student states[C]//*Proceedings of the 12th International Conference on Educational Data Mining*. Montreal: IEDM, 2019.
- [21] SINHA T, JERMANN P, LI N, et al. Your click decides your fate: Leveraging clickstream patterns in MOOC videos to infer students' information processing and attrition behavior[EB/OL]. (2014-09-16)[2021-03-01]. <https://arxiv.org/abs/1407.7131>.
- [22] YU C H, WU J, LIU A C. Predicting learning outcomes with MOOC clickstreams[J]. *Education Sciences*, 2014, 9(2): 104.
- [23] WEN Y, TIAN Y, WEN B, et al. Consideration of the local correlation of learning behavior to predict dropouts from MOOCs[J]. *Tsinghua Science and Technology*, 2020, 25(3): 336-347.
- [24] WHITEHILL J, WILLIAMS J J, LOPEZ G, et al. Beyond prediction: Towards automatic intervention in MOOC student stop-out[C]//*Proceedings of the 8th International Conference on Educational Data Mining*. Madrid: IEDM, 2015: 171-178.
- [25] FEI M, YEUNG D. Temporal models for predicting student dropout in massive open online courses[C]//*Proceedings of 2015 ICDM Workshop on Data Mining for Educational Assessment and Feedback*. Piscataway: IEEE Press, 2015: 256-263.
- [26] WANG W, YU H, MIAO C. Deep model for dropout prediction in MOOCs[C]//*Proceedings of the 2nd International Conference on Crowd Science and Engineering*. New York: Association for Computing Machinery, 2017: 26-32.
- [27] TANG C, OUYANG Y, RONG W, et al. Time series model for predicting dropout in massive open online courses[C]//*Proceedings of 19th International Conference on Artificial Intelligence in Education*. Berlin: Springer, 2018: 353-357.
- [28] XIONG F, ZOU K, LIU Z, et al. Predicting learning status in MOOCs using LSTM[C]//*Proceedings of the 2019 ACM Turing Celebration Conference-China*. New York: Association for Computing Machinery, 2019: 1-5.
- [29] YIN S, LEI L, WANG H, et al. Power of attention in MOOC dropout prediction[J]. *IEEE Access*, 2020, 8: 202993-203002.
- [30] SUN D, MAO Y, DU J, et al. Deep learning for dropout prediction in MOOCs[C]//*Proceedings of the 8th International Conference on Educational Innovation through Technology*. Piscataway: IEEE Press, 2019.
- [31] ZHENG Y, GAO Z, WANG Y, et al. MOOC dropout prediction using FWTS-CNN model based on fused feature weighting and time series[J]. *IEEE Access*, 2020, 8: 225324-225335.
- [32] MORENO-MARCOS P M, MUÑOZ-MERINO P J, MALDONADO-MAHAUAD J, et al. Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs[J]. *Computers & Education*, 2020, 145: 103728.
- [33] KLOFT M, STIEHLER F, ZHENG Z, et al. Predicting MOOC dropout over weeks using machine learning methods[C]//*Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*. Stroudsburg: ACL, 2014: 60-65.
- [34] NAGRECHA S, DILLON J Z, CHAWLA N V. MOOC dropout prediction: Lessons learned from making pipelines interpretable[C]//*Proceedings of the 26th International Conference on World Wide Web Companion*. New York: Association for Computing Machinery, 2017: 351-359.
- [35] LIANG J, LI C, ZHENG L. Machine learning application in MOOCs: Dropout prediction[C]//*Proceedings of 11th International Conference on Computer Science & Education*. Piscataway: IEEE Press, 2016: 52-57.
- [36] AL-SHABANDAR R, HUSSAIN A, LAWS A, et al. Machine learning approaches to predict learning outcomes in massive open online courses[C]//*Proceedings of 2017 International Joint Conference on Neural Networks*. Piscataway: IEEE Press, 2017: 713-720.
- [37] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. *Radiology*, 1982, 143(1): 29-36.

Behavior based MOOC user dropout predication framework

CHEN Hui^{1,*}, BAI Jun¹, YIN Chuantao², RONG Wenge¹, XIONG Zhang¹

(1. School of Computer Science and Engineering, Beihang University, Beijing 100191, China;

2. School of Sino-French Engineering, Beihang University, Beijing 100191, China)

Abstract: Though the massive open online courses (MOOC) have greatly changed the way of learning, properly understanding the user's behavior and then predication of dropout is one of the most challenging tasks. MOOC have significantly altered the way that people learn, yet one of the most difficult challenges is correctly interpreting user behavior and then predicting dropout. In this research, to improve the dropout prediction performance, we firstly analyzed users and courses from the perspective of activities by using the long short term memory mechanism. In this study, we used the long short term memory mechanism to analyze users and courses from the perspective of activities in order to improve the dropout prediction performance. Afterwards we further proposed a multi-attention based multi-perspective feature enhancement method to investigate the correlated activities among users and courses. Finally, we provided a gated mechanism-based feature integration framework for dropout prediction. The experiment study on the public dataset has shown our framework's promising potential, thereby making it possible to better investigate the reason beneath these phenomena and improve the overall study experience. The experiment study on the open dataset has demonstrated the promising potential of our framework, allowing us to more thoroughly explore the causes of these events and enhance the learning environment as a whole.

Keywords: massive open online courses; prediction framework; user; content; study behavior

Received: 2021-04-12; Accepted: 2021-04-21; Published Online: 2021-04-23 16:29

URL: kns.cnki.net/kcms/detail/11.2625.V.20210423.1555.002.html

Foundation item: National Natural Science Foundation of China (61977003)

* Corresponding author. E-mail: chenhui@buaa.edu.cn