

基于生成模型的学生在线学习表现预测混合方法研究

段超 王一晴 王洁 张明焱

浙江师范大学浙江省智能教育技术与应用重点实验室 浙江 金华 321004

(duanchao@zjnu.edu.cn)

摘要 学习表现预测利用在线学习平台的学生学习行为数据来识别存在学业风险的学生,可以帮助教师及时进行干预,然而该方式面临着数据不平衡问题,这使得准确识别存在学业风险的学生尤为困难。针对当前解决策略中变分自编码器(Variational Autoencoder,VAE)不能保证生成样本的合理性,生成对抗网络(Generative Adversarial Network,GAN)在处理时间序列数据时易引入新的错误,并且生成器和判别器任何一方训练得过于出色或不足都会导致生成数据质量下降等问题。提出了一种新的基于生成模型的学生在线学习表现预测混合方法。具体而言,首先利用融合双向长短期记忆网络(Bidirectional Long Short-term Memory,BiLSTM)的VAE对GAN进行初始化,不仅能从更加稳定的点开始训练,而且能更好地理解学生行为序列数据前后之间的关联关系和周期性特征;其次,判别器中引入多头注意力机制,增强其对真实数据和生成数据的区分能力,进而与生成器不断博弈;最后,将深度生成模型与经典重采样策略(Synthetic Minority Oversampling Technique,SMOTE)基于Blending集成学习的思想进行融合,有效结合数据和算法两个方面的优势,提高了模型整体的生成能力。在两个真实学生数据集上进行了大量实验,结果表明,该模型可以生成高质量的数据,从而提升预测模型对存在学业风险学生的识别能力,从第一单元开始,在4个评价指标上较基线方法均有提升。

关键词:学习表现预测;生成模型;双向长短期记忆网络;注意力机制;集成方法

中图分类号 TP391

Research on Hybrid Methods for Predicting Students' Online Learning Performance Based on Generative Model

DUAN Chao,WANG Yiqing,WANG Jie and ZHANG Mingyan

Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province,Zhejiang Normal University,Jinhua,Zhejiang 321004, China

Abstract Learning performance prediction can help teachers to intervene in time by using the learning behavior data of students on the online learning platform to identify at-risk students, but it faces the problem of data imbalance, which makes it particularly difficult to accurately identify at-risk students. Addressing the issues that the mainstream deep generative model VAE, cannot guarantee the rationality of generated samples in current solution strategies, and GAN tends to introduce new errors when processing time-series data, with either over-training or under-training of the generator or discriminator leading to a decline in the quality of generated data, this paper proposes a new prediction method for student learning performance based on generative models. Firstly, the VAE based on bidirectional long short-term memory(BiLSTM) is utilized to initialize the GAN, enabling it to start training from a more stable point while also better understanding the correlation and periodic characteristics between subsequent data points in the student behavior sequence. Secondly, a multi-head attention mechanism is introduced in the discriminator part to enhance its ability to distinguish between real data and generated data, and then continue to game with the generator. Finally, the deep generative model and the classical resampling strategy SMOTE are integrated based on the idea of Blending ensemble learning, which effectively combines the advantages of data and algorithm to improve the overall generation ability of the model. A large number of experimental results on real student data sets show that the model can generate high-quality data to improve the recognition ability of the prediction model for at-risk students, and is superior to the baseline method in multiple evaluation indicators.

Keywords Learning performance prediction, Generative model, Bidirectional long short-term memory, Attention mechanism, Ensemble method

基金项目:国家自然科学基金(62207027,62177024);教育部产学研合作协同育人项目(220906424035704);浙江省教育科学规划课题(2023SCG369)

This work was supported by the National Natural Science Foundation of China(62207027,62177024),University-Industry Collaborative Education Program(220906424035704) and Zhejiang Province Educational Science and Planning Research Project(2023SCG369).

通信作者:张明焱(mingyanzhang@zjnu.edu.cn)

1 引言

教育数据挖掘通过分析多源异构数据来支持与教学和学习相关的决策,受到了学术界和工业界的广泛关注。学生学习表现的早期预测作为教育数据挖掘领域重要的研究方向之一,通过构建预测模型在学期较早阶段预测学生的最终表现,以识别出尽可能多的存在学业风险的学生,方便教师尽早做出及时的干预措施,降低学生不及格或辍学的比例^[1]。

随着大数据和人工智能技术的发展,大量机器学习技术被引入学生学习表现领域,并取得了相当大的成就。特别是深度学习技术凭借其多层神经网络结构的独特优势,在自然语言处理、语音识别、图像处理等多个领域不断取得突破性的研究成果,这一系列成就进一步激发了教育相关研究者对深度学习技术的深入探索,一些学者已经在学生学习表现预测的研究工作中取得了一定的进展。Patil 等^[2]探讨了递归神经网络在预测学生学习表现方面的应用,包括不同的递归神经网络架构与其他机器学习技术对比的结果。而 Kim 等^[3]则专门针对在线课程中的学习表现预测提出了一种新的深度学习 GritNet,该方法基于 BiLSTM 并将学习表现预测问题重新定义为一个序列事件预测问题。Yao 等^[4]则针对长行为序列建模问题,提出了一个端到端的双层自注意力网络 DEAN,用于对不同专业的学生成绩进行预测。但大量研究多采用单一的深度学习模型,制约了预测性能。

虽然现有研究取得了一定的进展,但是学生学习表现的早期预测面临着教育数据不平衡问题的挑战。因为一般情况下,大部分学生会努力学习以通过课程考核,只有少数学生会出现不及格或者是辍学的现象。那么,对于一个风险率为 3% 的不平衡数据集来说,如果一个模型简单地预测所有学生都是成功学生,那么该模型也可以达到 0.97 的准确率。然而,该模型未能识别出任何有学业风险的学生。针对此问题,研究者们相继从数据和算法两个层面提出了解决方案。算法方面的最新方法主要基于集成学习,数据方面方法的应用更为广泛。最为热门的合成少数类过采样 SMOTE 方法^[5]通过在特征空间中生成少数类样本的合成实例来增加少数类的样本数量,改善模型对少数类的识别能力,但是存在严重的类重叠问题。随着深度生成模型不断取得技术突破,深度生成模型因其强大的样本生成能力被逐渐应用于不平衡数据研究中。Du 等^[6]提出了一种基于潜在变分自编码器 VAE 和深度神经网络 DNN 的集成框架 LVAE, LVAE 旨在了解学业风险学生的潜在分布并生成风险学生样本以获得平衡的数据集。Sarwat 等^[7]提出了 CGAN-DSVM 算法,用于预测学生在校和家庭辅导支持性学习下的表现,通过改进的条件生成对抗网络 (Conditional Generative Adversarial Network, CGAN) 生成新的训练数据。虽然这些研究工作在学生学习表现预测上取得了一定的研究结果,但是仍然存在着主流的深度生成模型 VAE,并不能保证生成样本的合理性,GAN 模型在处理时间序列数据时易引入新的错误且生成器和判别器之间的平衡并非总是容易实现,而任何一方训练的过于出色或不足都会导致生成数据质量下降等问题。以上研究未考虑结合模型和数据两个方面进行学习表现预测。

综上所述,为有效缓解上述问题带来的挑战,本研究提出了一种新的基于生成模型的学生在线学习表现预测混合方法

(BiLSTM and Multi-head Attention Based VAE-GAN Blend with Smote, BMVGS)。该方法引入了 BiLSTM 和多头注意力机制的 VAE-GAN 生成模型与 SMOTE 经典过采样方法基于 Blending 集成学习的思想进行融合。本研究主要工作包括以下 3 个方面:1) 引入基于 BiLSTM 的 VAE 对 GAN 模型进行初始化,使得 GAN 模型能够从一个更加稳定的点开始训练的同时,也能够很好地理解学生学习行为的前后关联性和周期性特征;2) 将多头注意力机制引入判别器部分,使得生成器和判别器在训练过程中保持相互依存与持续优化的关系;3) 将深度生成模型和经典重采样策略 SMOTE 基于 Blending 集成学习的思想进行融合,充分发挥数据和模型两方面的各自优势,提高了 BMVGS 模型整体的生成能力和预测效果。

2 相关工作

2.1 基于深度学习的学习表现预测

随着机器学习技术的不断发展,学者们逐渐将研究目标从机器学习过渡到了深度学习。Zheng 等^[8]提出了一种辍学预测深度融合模型,首先利用 CNN 获取局部特征表示并利用自注意力机制学习不同特征表示之间的相关性,然后利用 BiLSTM 获得时间序列特征向量表示。Zhang 等^[9]提出了一种通过深度学习模型对学生的微观行为模式进行分析以实现学习成绩早期预警的 LAA 方法,它首先利用 LSTM-autoencoder 对学生行为的时间序列数据进行去噪和特征提取,然后通过注意力机制计算影响权重并识别出 at-risk 学生和最早干预时间点。Chen 等^[10]则构建了一种基于多头自注意力机制的学生成绩预测模型,以自动选择更重要的高阶行为组合特征,提高模型的预测精度。Huang 等^[11]通过对学习模式和学习动机的深入探讨,提出一种基于行为演化规律的学习模式识别及效果预测方法。Wang 等^[12]针对学生生成内容中的噪声和现有的基于无符号图模型的局限性,将有符号图神经网络与大型语言模型进行融合,提出了 LLM-SBCL 模型,提高了对学生表现的多模态预测性能。

已有研究在模型优化方面做出了突出的贡献,提高了在线学习表现预测的精度,然而以上研究单从模型层面进行优化,忽略了数据方面的优化问题。众所周知,数据本身也会对成绩预测的实验结果产生重要影响,因而本研究期望从数据和模型两个方面同时对结果进行优化。

2.2 不平衡数据处理策略

针对教育数据不平衡问题,研究人员从数据和算法两个层面展开了深入而细致的研究。根据在线教育领域的已有研究成果,可以将数据级方法分为过采样和深度学习方法两大类。其中,深度学习方法中最为主流的深度生成模型包括变分自编码器 VAE^[13]和生成对抗网络 GAN 模型^[14]。

过采样是一种通过生成新实例来增加少数类实例规模以获得平衡类的方法。Rachburee 等^[15]基于 SMOTE 的相似原理,在 463 956 条学生数据记录中比较了使用 SMOTE, SVM-SMOTE 和 ADASYN 来解决不平衡类以预测学生表现的过采样方法。Rahim 等^[16]则比较了 SMOTE, Borderline SMOTE 和 ADASYN 这 3 种过采样技术在学生辍学预测中的效果。结果表明,ADASYN 在提高学生辍学预测的召回率方面优于其他两种方法。

变分自编码器 VAE 通过引入概率编码和解码的方式,学习输入数据的潜在低维表示,并能够通过潜在空间中的采样生成新的数据样本。Zhang 等^[17]介绍了一种基于图神经网络的矩阵填充模型(Graph-VAE),用于预测学生分数。该模型能够自动提取特征并更好地描述学生与课程之间的相关性和差异。

生成对抗网络 GAN 通过生成器和判别器两个神经网络的相互竞争和对抗训练,使生成器能够生成与真实数据难以区分的样本。Chui 等^[18]提出了 ICGAN-DSVM 算法,用于预测学生在学校和家庭辅导支持性学习下的表现。随后针对相同领域的相同研究问题,Sathiyapriya 等^[19]提出了 ICGAN-PSVM 算法,首先使用改进的 MCSO 算法从大学数据集中选择合适的特征,然后使用所提方法进行预测。与前两项研究不同,Waheed 等^[20]提出了一种基于对抗的新型方法(SC-GAN),用于在教育环境中对顺序数据进行上采样以消除学生学业成绩的不平衡。以上方法相对单一,难以进一步提高预测性能。

算法层面的方法侧重于集成学习,通过集成多个基学习器来缓解不平衡给分类器带来的影响,常见的集成方法有 Bagging, Boosting, Stacking 和 Blending 等。Lenin 等^[21]针对收集到的学生数据高度不平衡问题,探索了随机森林和 Ada-Boost 等几种集成机器学习技术,试图开发一个模型来预测学生在 Meghalaya 大学的表现。Rabelo 等^[22]则基于对学生辍学问题的表征和知识发现过程的应用,提出了一种结合逻辑回归、神经网络和决策树的集成模型来改进高等教育学生

辍学预测。鉴于集成学习模型的良好表现,本研究充分利用各方面已有研究成果并进行进一步优化,通过集成学习充分考虑数据和算法的优势,提出了 BMVGS 模型。

基于以上两方面相关工作的研究综述可知,“学习表现”是学生在在学习过程中所展现出的学业成绩、学习行为、学习状态、社会互动等多方面特征的综合体现。在研究中,可以根据具体目标选择某一维度或多维度进行量化分析,并且大量的已有研究表明,在学习表现预测中可以重点关注学生的学业成绩和学习行为。

3 BMVGS 模型

3.1 模型整体架构

BMVGS 模型的架构如图 1 所示,模型框架主要包括四大部分,分别为输入层、基学习器层、元学习器层和输出层。具体来说,首先是输入层,采用留出法划分原始数据集;其次是基学习器层,VAE-GAN 模型的编码网络负责将输入的真实样本转换为隐变量,随后被用作生成网络的输入以生成新的数据样本,判别网络则评估这些生成样本的真实性,然后将生成数据 R_G 与训练集组合成新的平衡训练集输入到分类器随机森林(Random Forest, RF)中进行训练,SMOTE 输出的 R_G' 则直接输入 RF 中进行训练,将验证集和原始测试集输入到训练好的两个基模型中,输出的结果进行拼接进而形成元训练集和元测试集;再次是元学习器层,它利用元训练集和元测试集进行元学习器逻辑回归 LR 的训练和预测;最后是输出层,将最终预测的分类结果进行输出。

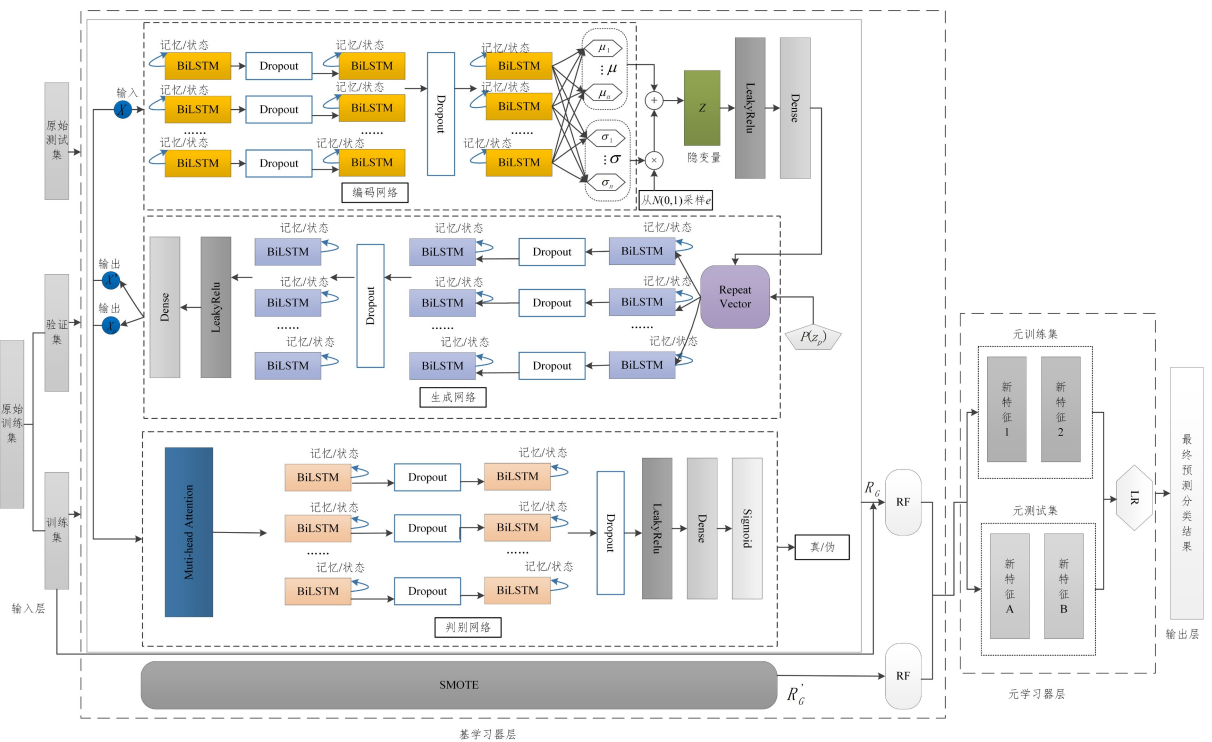


图 1 BMVGS 模型架构

Fig. 1 Architecture of BMVGS model

3.2 输入层

输入层的作用是首先将原始数据集采用留出法按照一定的比例规则划分为原始训练集和原始测试集,然后根据 Blending 算法的原理将原始训练集再次按照一定的数据比例划分成训练集和验证集,用于基学习器层模型的训练、验证和预测。

3.3 基学习器层

基学习器层包含两个基模型,分别是 VAE-GAN-RF 模型和 SMOTE-RF 模型。其中,在 VAE-GAN 模型中引入了 BiLSTM 以更加深入地理解学业风险学生行为序列数据中所蕴含的潜在信息和周期性特征,在其判别网络部分则引入了

多头注意力机制以更有效地聚焦于输入数据的关键区域并增强判别网络对真实样本和生成样本的区分能力。当判别网络训练得更好时,又会与生成网络继续博弈,从而共同进步。因此,本部分对基于 BiLSTM 和多头注意力机制的深度生成模型 VAE-GAN 进行详细分析。

1) 编码网络

给定输入 x , 数据形状大小为 (S, U, F) , S 为每次训练所选取的样本数, U 与所取的学习单元相关, 即可取 $1, 2, \dots, 12$ 。 F 则为每单元所包含的学生行为特征数, 共 9 个维度。 $h_{-E}^{\wedge}(u)$ 则为 t_u 时刻编码器的隐藏状态, 其中, $u \in \{1, 2, \dots, U\}$ 。 则由 BiLSTM 网络构成的编码网络部分对学业风险学生的在线学习行为序列编码过程表示为:

$$\overrightarrow{h}_E^{(u)} = \overrightarrow{LSTM}^E(x_u, \overrightarrow{h}_E^{(u-1)}) \quad (1)$$

$$\overleftarrow{h}_E^{(u)} = \overleftarrow{LSTM}^E(x_u, \overleftarrow{h}_E^{(u+1)}) \quad (2)$$

其中, $\overrightarrow{h}_E^{(u)}$ 表示时间步为 u 时, 前向 LSTM 的隐藏层状态; \overrightarrow{LSTM} 表示前向 LSTM 单元; x_u 是时间步 u 的输入; $\overleftarrow{h}_E^{(u)}$ 表示时间步为 $u-1$ 时, 正向 LSTM 的隐藏层状态。后向 LSTM 公式中各向量所表示的含义则亦然。

那么, 编码网络最终的状态可表示为:

$$h_E^{(U)} = \text{concat}(\overrightarrow{h}_E^{(U)}, \overleftarrow{h}_E^{(U)}) \quad (3)$$

其中, concat 表示向量拼接。

由最终状态所产生的学生潜在特征向量 z 可以表示为:

$$z = h_E^{(U)} \quad (4)$$

在向量表征框架下, 各变量的分布特征隐藏着学业风险学生的潜在属性, 这些属性既能揭示学生的内在特质, 也能映照出其学习历程及当前的学习状况。模型的目标在于生成与数据集中样本共享相同潜在特性的新样本, 因此编码网络的优化目标是 minimized KL 散度, 即期望 $N(0, 1)$ 与潜在变量分布越相似越好。其损失函数如下:

$$L_E = KL(N(\mu, \sigma^2) \| N(0, 1)) \\ = -\frac{1}{2} \sum_{j=1}^D (1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2) \quad (5)$$

其中, μ_j 和 σ_j 分别是潜在变量第 j 维的均值和标准差, D 是潜在变量的维度。

2) 生成网络

VAE-GAN 模型中的生成网络与 VAE 的解码网络共享权重, 任务是将编码网络生成的潜在表示 z 重新构造出原始的时间序列数据。与编码网络结构类似, 生成/解码网络包含 3 层 BiLSTM 网络, 每一层的前向 LSTM 对于每个时间步 u , 其计算式为:

$$\overrightarrow{h}_D^{(u)} = \overrightarrow{LSTM}^D(z_u, \overrightarrow{h}_D^{(u-1)}) \quad (6)$$

其中, z_u 是时间步 u 的输入特征向量, 形状为 (S, L) , 其中 L 为潜在空间的维度。

后向 LSTM 对于每个时间步 u , 其计算式为:

$$\overleftarrow{h}_D^{(u)} = \overleftarrow{LSTM}^D(z_{U-u}, \overleftarrow{h}_D^{(u+1)}) \quad (7)$$

其中, z_{U-u} 为时间步 $U-u$ 的输入特征向量, 形状为 (S, L) 。

最终的输出为前向和反向 LSTM 隐藏状态的拼接, 即:

$$h_D^{(U)} = \text{concat}(\overrightarrow{h}_D^{(U)}, \overleftarrow{h}_D^{(U)}) \quad (8)$$

其输出形状为 $(S, L, 2 * 4)$ 。

生成/解码网络的目标不仅要使重构的时间序列数据最小, 还要尽可能地骗过判别器, 因此其损失函数可表示为:

$$L_G = E_{z_p \sim P_{z_p}} [\log(1 - D(G(z_p)))] + E_{x \sim P_{\text{data}}(x), z \sim P_z(z)}$$

$$\gamma(x - G(z))^2 \quad (9)$$

其中, x 为输入的真实样本; $G(z_p)$ 表示 z_p 输入到网络所生成的数据样本; $D(\cdot)$ 为对输入的样本数据进行打分操作, 以判断其真实性; γ 为引入的超参数, 以平衡最初的生成网络和解码网络之间的误差。

3) 判别网络

判别器的输入是真实学业风险学生时间序列数据 x 、生成的数据 x' 以及 x_p , 其任务是区分真实数据和生成数据。在输入层之后引入多头注意力机制, 允许模型在不同的表示子空间中关注不同的信息以提高判别器的特征提取和区分能力。在本研究中, 多头注意力机制的头数设置为 3, 以生成的数据 x' 作为输入层数据为例, 输入序列 $x_u^G = [x_{u,1}^G, x_{u,2}^G, \dots, x_{u,K}^G]$ 为生成的学业风险学生特征, W_h^Q, W_h^K, W_h^V 为 3 个可学习的矩阵参数, 用于 Q_h, K_h, V_h 矩阵的生成。在得到矩阵之后, 开始对注意力权重矩阵进行计算。

其中, K_h^T 是 K_h ($h = 0, 1, 2$) 的转置矩阵, d_{head_h} 为 9, $\sqrt{d_{head_h}}$ 的值为 3, 其目的是为了进行归一化操作, 然后对结果的每一行使用 softmax 函数进行计算。

$$\text{softmax}(g_a) = e^{g_a} / \sum_{b=1}^n e^{g_b} \quad (10)$$

其中, g_a 是 $Q_h \times K_h^T / 3$ 矩阵一行内第 a 列的数值, g_b 则为 $Q_h \times K_h^T / 3$ 矩阵一行内第 b 列的数值, n 是 $Q_h \times K_h^T / 3$ 的列数。

将得到的注意力权重矩阵与输入序列 $x_u^G = [x_{u,1}^G, x_{u,2}^G, \dots, x_{u,K}^G]$ 进行拼接, 可得到多头注意力机制模块的输出序列, 即 $l_u^M = [l_{u,1}^M, l_{u,2}^M, \dots, l_{u,K}^M]$ 。

而后, 将由输入数据经过多头注意力机制处理后得到的输出经过两层 BiLSTM 网络, 得到形状为 $(S, 2 * 16)$ 的 h_U 输出。在每个时间步 u 的输出 h_u 上应用 10% 的 dropout, 即随机将 10% 的元素置为 0 以防止过拟合。应用 Leaky ReLU 激活函数, 引入非线性, 帮助模型学习更复杂的特征, 然后应用全连接层, 输出维度为 1。最后应用 Sigmoid 激活函数, 将输出压缩至 0 到 1 之间。其损失函数为:

$$L_D = \log(\text{Dis}(x)) + \log(1 - \text{Dis}(\text{Dec}(z))) + \log(1 - \text{Dis}(\text{Dec}(\text{Enc}(x)))) \quad (11)$$

即来自真实样本数据、从先验分布中采样生成的数据以及由编码器重构的数据三者损失之和。

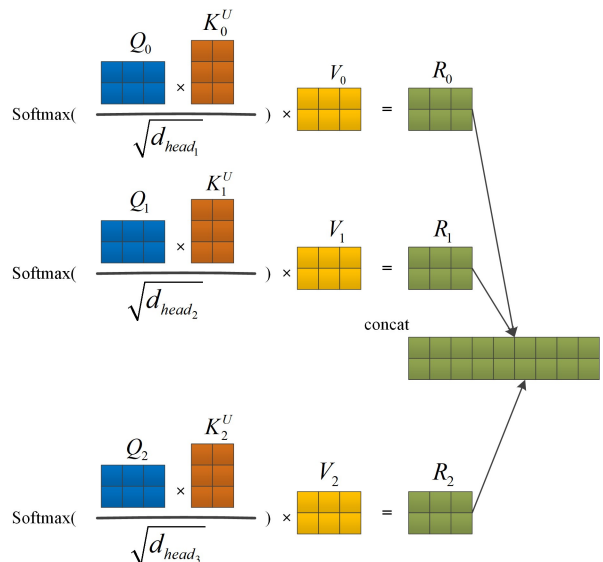


图 2 头注意力机制得分矩阵计算

Fig. 2 Multi-head-attention score matrix calculation

3.4 元学习器层

在 BMVGS 模型元学习器中,选择逻辑回归(Logistic Regression, LR)模型为元学习器进行模型最终的分类型预测,将模型记作 \mathfrak{L} 。LR 算法的复杂度较低,降低了过拟合风险。对于元学习器而言,其输入训练集为上一层基学习器层输出的验证集预测结果的拼接,记为 L_T 。测试集则为训练好的基学习器对原始测试集预测结果的拼接,记为 L_i 。则元学习器的训练过程可以表示为:

$$\mathfrak{L} = \mathcal{L}_{LR}(L_T) \quad (12)$$

其中, $\mathcal{L}_{LR}(\cdot)$ 表示逻辑回归 LR 算法。

3.5 输出层

将元测试集 L_i 输入训练完成的元学习器 \mathfrak{L} 中以得到最终的分类型结果,预测结果依然为标志着学生能否通过课程考核的二分类标签 $y, y \in \{0, 1\}$ 。

4 实验结果与分析

4.1 数据集及评价指标

1) 数据集

在这项研究中,实验数据收集自美国一所学校完全在线的课程,模型性能对比的数据为 2021 年秋季学期学生在线学习行为日志数据,该数据集共包含 16525 名学生。对学生原始在线学习日志数据进行清洗和预处理后,将学生的用户名和系统分配的用户编号结合起来形成一个唯一的 ID 来标识学生,所有学生数据按照单元划分为 0-10 单元,还包含 1 个期中考试单元,共 12 个单元。每个单元包含:Assignment 查看和提交作业行为次数;TestQuiz 参与测验行为次数以及 DiscussionBoard 讨论行为次数等 9 个在线学习行为相关的变量。

学生的学业成绩被描述为一个类别型变量,它显示了某位学生是否能够通过课程考核。就学生的学业成绩数据而言,设置存在学业风险的成绩阈值,低于阈值的学生被标记为“1”,高于阈值的学生被标记为“0”。按照阈值划分,数据集中包含 4362 个学业风险学生样本数据和 12163 个学业成功学生样本数据,存在数据不平衡问题。

同时,为验证本研究设计方法的普遍性和适用性,将训练好的模型在 2022 年春季学期数据集上进行测试。该数据集包含 12157 人,其中学业风险学生有 3468 人,学业成功学生有 8689 人。同样共 12 个学习单元,每单元包含 9 个学习行为变量。

本研究采用双数据集实验设计,分别用于模型性能对比和泛化能力测试。数据集采集自真实教育场景中的学生学习时间序列数据,全面记录了学生在不同时间点上的学习行为、表现和状态等信息,能够有效表征学习过程的动态演变特征。基于这些时序数据,本研究构建了 BMVGS 模型,通过提取学生的潜在特征向量,生成具有高度真实性的学业风险学生样本数据,在实现教育数据集类别平衡的同时,也为学生学习表现的精准预测提供了可靠的数据基础和技术支撑。

2) 评价指标

在本研究中,使用 Accuracy, Precision, Recall 和 F1 值评估模型性能。准确率 Accuracy 是展示一个模型性能最直观的评价指标。召回率 Recall 是模型捕获的正样本,也就是存在学业风险学生的比率;精度 Precision 指被预测为学生风险学生的学生中实际上是风险学生的比率。在许多分类任务

中,高精度和高召回率是不能同时实现的。因此,F1 被作为精度和召回率的调和均值。一般来说,F1 值越高,模型的预测性能越好。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (16)$$

其中,TP 表示该学生为学业风险学生并且模型正确预测为学业风险学生;TN 表示该学生为学业成功学生并且模型正确预测为学业成功学生;FP 表示将学业成功的学生误判为学业风险学生;FN 则表示将学业风险学生误判为学业成功学生。

4.2 基线模型

为了验证本研究所设计方法的有效性,将 BMVGS 模型与以下 7 种模型进行对比,这 7 种模型均使用随机森林 RF 进行预测。

Origin:使用没有数据平衡方法的 RF 分类器得到初始基线结果,验证数据平衡方法是否有利于预测性能的提高。

SMOTE^[5]:是一种用于处理不平衡数据集的过采样技术,通过在少数类样本之间生成合成样本来增加少数类的代表性,从而改善分类器对少数类的识别能力。

VAE^[13]:是一种基于深度学习的生成模型,通过最小化数据的证据下界(ELBO),利用变分推断来近似后验分布,从而实现了对数据的有效学习和生成。

GAN^[14]:是一种基于深度学习的无监督学习算法,通过训练一个生成器和一个判别器进行对抗学习,估计数据样本的潜在分布并生成新的数据样本。

VAE-GAN^[23]:是一种结合了变分自编码器(VAE)和生成对抗网络(GAN)深度学习模型,通过融合两者优点来提高样本生成的质量和效率。

RNN-VAE-GAN^[24]:采用循环变分自编码器(VAE)作为生成器,将双向 RNN 作为鉴别器。循环 VAE 将时间动态捕获到学习的随时间变化的潜在空间上,而对抗性训练鼓励模型生产真实的时间序列数据。

LSTM-VAE-GAN^[25]:在 VAE-GAN 模型中引入 LSTM 形成 LSTM-VAE-GAN 模型,提升时间序列异常检测的性能。

4.3 实验设置

本次实验研究是在 64 位 Windows 11 操作系统的集成开发平台 Jupyter Notebook 6.5.4 上进行,使用 Python 3.9.17 编程语言开发,利用 Tensorflow 2.10.0 深度学习框架,所使用的主要算法库有 sklearn 库、numpy 库、pandas 库、imblearn 库、matplotlib 库等。硬件环境为 Intel^(R) Core^(TM) i7-4790K CPU @ 4.00 GHz、NVIDIA GeForce GTX 1080。在数据集划分方面,将数据集按照 8:1:1(训练集:验证集:测试集)的规则进行划分,BMVGS 模型的优化器使用 Adam,学习率设置为 0.0002,批大小 batch size 的值设置为 512,迭代次数 epoch 为 20。其次,设置 SMOTE 方法的参数 sampling_strategy 为 minority 策略。

4.4 性能对比与泛化测试

4.4.1 模型性能对比

为了全面评估 BMVGS 模型的性能,将其与多种基线方法置于相同的实验条件下进行了对比分析。针对学业风险学

生的早期预测任务,各模型的实验结果如表 1 所列。鉴于篇幅所限,仅展示了第 1 单元、第 3 单元等关键学习阶段的预测结果。其中,最佳性能结果用黑体加粗,次优结果用下划线标记。

表 1 学业风险学生早期预测结果比较
Table 1 Comparison of early prediction results of academic risk students

评估指标	模型	第 1 单元	第 3 单元	第 4 单元	第 5 单元	第 7 单元	第 9 单元
Accuracy	Origin	0.779	0.831	0.846	0.861	0.903	0.889
	SMOTE	0.752	0.881	0.885	0.902	0.909	0.901
	VAE	0.807	0.910	0.922	0.925	0.931	0.929
	GAN	0.808	0.929	0.930	0.935	0.936	0.937
	VAE-GAN	0.823	0.932	0.934	0.942	0.946	0.943
	RNN-VAE-GAN	0.843	0.934	0.935	0.945	0.949	0.948
	LSTM-VAE-GAN	<u>0.851</u>	<u>0.935</u>	<u>0.937</u>	<u>0.947</u>	<u>0.950</u>	<u>0.950</u>
	BMVGS	0.875	0.944	0.950	0.953	0.956	0.956
Precision	Origin	0.637	0.796	0.774	0.796	0.876	0.824
	SMOTE	0.565	0.754	0.764	0.792	0.830	0.800
	VAE	0.777	0.881	0.894	0.901	0.909	0.909
	GAN	0.811	0.914	0.923	0.922	0.920	0.917
	VAE-GAN	0.822	0.919	0.924	0.932	0.922	0.925
	RNN-VAE-GAN	0.850	0.920	0.923	0.924	0.924	0.917
	LSTM-VAE-GAN	<u>0.850</u>	<u>0.921</u>	<u>0.926</u>	<u>0.932</u>	<u>0.925</u>	<u>0.928</u>
	BMVGS	0.873	0.928	0.936	0.941	0.933	0.960
Recall	Origin	0.248	0.474	0.629	0.675	0.738	0.770
	SMOTE	<u>0.597</u>	<u>0.829</u>	<u>0.849</u>	0.853	0.833	0.844
	VAE	0.379	0.763	0.802	0.803	0.823	0.815
	GAN	0.359	0.807	0.803	0.824	0.831	0.838
	VAE-GAN	0.431	0.814	0.817	0.842	0.869	0.854
	RNN-VAE-GAN	0.492	0.819	0.822	<u>0.863</u>	0.880	<u>0.884</u>
	LSTM-VAE-GAN	0.526	0.824	0.829	0.862	<u>0.883</u>	0.878
	BMVGS	0.615	0.853	0.869	0.876	0.897	0.871
F1	Origin	0.338	0.583	0.639	0.718	0.798	0.784
	SMOTE	0.554	0.788	0.798	0.820	0.829	0.818
	VAE	0.508	0.818	0.845	0.849	0.864	0.859
	GAN	0.495	0.857	0.859	0.824	0.873	0.875
	VAE-GAN	0.561	0.863	0.867	0.885	0.895	0.888
	RNN-VAE-GAN	0.623	0.867	0.870	0.893	0.901	0.900
	LSTM-VAE-GAN	<u>0.650</u>	<u>0.870</u>	<u>0.875</u>	<u>0.896</u>	<u>0.904</u>	<u>0.902</u>
	BMVGS	0.721	0.889	0.901	0.907	0.915	0.913

通过表 1 所呈现的实验结果,可以观察到:

首先,所提出的 BMVGS 模型在所有评估指标上均展现出了最优的性能,尤其是在刚开始的几个学习单元上,相较于 Origin 模型,对学业风险学生的识别能力差距尤为明显。随着学习单元的递进,模型的准确率 Accuracy 呈现出稳步上升的趋势,这很可能归因于模型在处理更多训练数据的过程中,能够逐步捕捉到更加广泛且全面的特征表征。

其次,进入第 3 学习单元后,BMVGS 模型的召回率 Recall 就已达到 85.3%,精确度 Precision 也达到 92.8%,但是本实验选择第 4 学习单元作为最佳干预时间点。原因在于,在第 4 学习单元,模型的召回率 Recall 攀升至 86.9%,精确度 Precision 更是达到了 93.6% 的高水平,更为关键的是,自此单元之后,模型的识别能力趋于稳定,显示出良好的鲁棒性和可靠性。

最后,整体上居性能第二的是 LSTM-VAE-GAN 模型,这一结果并不令人意外。一方面,LSTM 网络可以有效地建模学生的学习轨迹,另一方面,LSTM-VAE-GAN 模型巧妙地

融合了主流的深度生成模型 VAE 与 GAN 的优势,在一定程度上实现了两者功能的优势互补。而就召回率 Recall 来说,LSTM-VAE-GAN 模型并没有表现出良好的学业风险学生识别性能,相反,SMOTE 方法在多数学习单元中展现出了较为出色的识别能力。然而,值得注意的是,SMOTE 采样方法在提升召回率的同时,也伴随着较高的误分辨率问题,这导致其在精确度 Precision 方面通常不及深度生成模型。

4.4.2 模型泛化测试

为了验证所提模型的泛化能力,将 BMVGS 模型在 2022 年春季学期数据集上进行测试。由于已经确定第 4 学习单元为最佳干预时间点,因此利用该数据集中前 4 个学习单元的学生在线学习行为数据,对本学期可能处于学业风险的学生进行预测。实验结果如表 2 所列,最佳性能结果用黑体加粗表示,次优结果用下划线标记。

从表 2 的实验结果中,可以得到如下结论:

首先,所提 BMVGS 模型在各项指标上仍然具有最佳的性能。在 2022 年数据集的第 4 学习单元上仍然具有 90% 以

上的精确识别率以及高达约 80% 的召回率,这些都表明了 BMVGS 模型具有良好的泛化能力。在此数据集上,仍然可以将第 4 学习单元作为干预的最佳时机。

其次,相较于 Origin 模型,BMVGS 模型在各学习单元的各项指标上均实现了显著提升。如在最佳干预时间点第 4 学习单元上,BMVGS 模型成功多识别出约 30% 的学业风险学生,即 Recall 值得到提升。这一结果有力地证明了数据平衡的重要性,同时也验证了本文方法在增强模型预测能力方面的有效性。这也意味着,如果将 BMVGS 模型集成到在线学习平台亦或学习表现早期预警系统中,那么基于模型的早期预测结果,教师可以在最佳干预时间点对这些多识别出的

30% 存在学业风险的学生进行针对性的辅导,若其中 25% 学生的学习表现显著提升,那么干预成功率则达到 83.3%。

再次,SMOTE 方法在应用于 2022 年数据集的不同学习单元时,依然展现出了较好的识别效能。具体而言,它在 Recall 召回率这一指标上的表现仅次于所提出的 BMVGS 模型。然而,在评估识别学生精确度的 Precision 精确率指标上,SMOTE 方法的表现仍然不及其他基线深度生成模型。

最后,从整体性能来看,LSTM-VAE-GAN 模型依旧占据第二的位置,其准确率 Accuracy 和综合评价指标 F1 值仅次于 BMVGS 模型。其在精确度 Precision 上与 RNN-VAE-GAN 模型对学业风险学生预测准确度的影响不分伯仲。

表 2 基于 2022 年数据集的泛化测试

Table 2 Generalization testing based on the 2022 dataset

评估指标	模型	第 1 单元	第 3 单元	第 4 单元	第 5 单元	第 7 单元	第 9 单元
Accuracy	Origin	0.756	0.793	0.794	0.809	0.814	0.837
	SMOTE	0.774	0.849	0.860	0.876	0.897	0.900
	VAE	0.795	0.865	0.885	0.891	0.904	0.910
	GAN	0.786	0.868	0.881	0.889	0.901	0.904
	VAE-GAN	0.817	0.891	0.899	0.902	0.916	0.922
	RNN-VAE-GAN	0.844	0.901	0.914	0.915	0.919	0.926
	LSTM-VAE-GAN	<u>0.868</u>	<u>0.906</u>	<u>0.916</u>	<u>0.920</u>	<u>0.924</u>	<u>0.928</u>
	BMVGS	0.881	0.921	0.927	0.932	0.938	0.945
Precision	Origin	0.587	0.614	0.591	0.603	0.769	0.799
	SMOTE	0.597	0.728	0.724	0.772	0.803	0.814
	VAE	0.720	0.852	0.900	0.901	0.913	0.929
	GAN	0.790	0.896	0.908	0.910	0.911	0.922
	VAE-GAN	0.755	0.891	0.907	0.912	0.915	0.929
	RNN-VAE-GAN	0.820	<u>0.908</u>	<u>0.918</u>	0.919	0.923	<u>0.932</u>
	LSTM-VAE-GAN	<u>0.868</u>	0.900	0.910	<u>0.921</u>	<u>0.924</u>	0.929
	BMVGS	0.902	0.919	0.926	0.931	0.938	0.951
Recall	Origin	0.242	0.446	0.492	0.526	0.586	0.592
	SMOTE	<u>0.629</u>	<u>0.747</u>	<u>0.784</u>	<u>0.796</u>	<u>0.824</u>	<u>0.828</u>
	VAE	0.433	0.617	0.657	0.681	0.719	0.728
	GAN	0.318	0.593	0.636	0.662	0.710	0.712
	VAE-GAN	0.606	0.689	0.708	0.712	0.767	0.783
	RNN-VAE-GAN	0.522	0.713	0.756	0.760	0.774	0.789
	LSTM-VAE-GAN	0.615	0.741	0.774	0.777	0.788	0.801
	BMVGS	0.637	0.781	0.798	0.816	0.830	0.844
F ₁	Origin	0.323	0.476	0.507	0.546	0.621	0.663
	SMOTE	0.610	0.734	0.760	0.781	0.817	0.821
	VAE	0.536	0.715	0.759	0.776	0.805	0.816
	GAN	0.447	0.713	0.748	0.767	0.798	0.804
	VAE-GAN	0.606	0.776	0.795	0.800	0.835	0.848
	RNN-VAE-GAN	0.638	0.799	0.829	0.832	0.842	0.855
	LSTM-VAE-GAN	<u>0.720</u>	<u>0.813</u>	<u>0.836</u>	<u>0.843</u>	<u>0.850</u>	<u>0.861</u>
	BMVGS	0.747	0.844	0.858	0.869	0.881	0.895

4.5 消融实验

为了验证 BMVGS 模型中核心组件对最终结果的重要贡献,本节在不平衡数据集上针对 BiLSTM 和多头注意力机制进行了消融实验,同样使用 4 个指标作为衡量模型性能的标准,以确保评估的全面性和准确性。具体而言,消融实验的设置依据所移除的组件分为以下两种情况:1) $BMVGS_{w/oBi}$ 它表示禁用 BiLSTM,仅依赖单向 LSTM 网络对学业风险学生时间序列数据进行学习和生成;2) $BMVGS_{w/oAttention}$ 表示禁用多头注意力机制,在判别器部分移除多头注意力层。基于 4.4 节实验结果的指导,已确定第 4 学习单元为最佳干预时间点,因此本小节专注于前 4 学习单元的数据进行模型的消融实验,实验结果如图 3 所示,该图直观地展示了这两种配置下模型性能的对比情况。

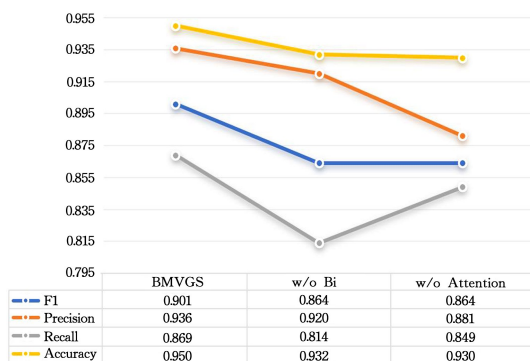


图 3 BiLSTM 和多头注意力机制对实验结果的影响

Fig. 3 Influence of BiLSTM and multi-head attention mechanism on the experimental result

通过图 3 所示实验结果的对比,可以得到以下结论:

1)与所提出的 BMVGS 模型相比,禁用 BiLSTM 而使用 LSTM 网络时,模型的性能明显有所下降,这一现象有力地验证了 LSTM 在捕捉学业风险学生潜在特征信息方面的局限性。相比之下,通过引入 BiLSTM 机制,模型能够更为全面地解析学生学习行为数据中蕴含的复杂关联性和周期性特征。

2)禁用多头注意力机制也得到了一个较差的结果,这是由于判别器在缺乏该机制的情况下,仅能捕获输入数据的单一维度信息,从而限制了其有效区分真实数据与生成数据的能力。而加入多头注意力机制后,判别器可以从多层面捕捉到更丰富的学生特征信息并且提高模型对关键特征的敏感度和理解,进而降低对噪声的敏感性的同时提高对真实数据和生成数据的区分能力。

3)禁用 BiLSTM 和禁用多头注意力机制对模型最终性能的影响相当接近,具体表现在准确率 Accuracy 和综合评价指标 F1 上几乎无显著差异。这一观察结果进一步印证了生成器与判别器在训练过程中是相互依存与持续优化的关系,任何一方训练得过于出色或者不足都会影响最终生成的数据样本质量。而禁用 BiLSTM 比禁用多头注意力机制的精确度 Precision 略高,这可能是由于模型此时更多地依赖于当前输入或局部特征进行决策,这有助于减少冗余信息的干扰,使模型在学习正样本即学业风险学生数据时更加专注和准确,从而相对提高了精确度。但同时又由于其不能更加全面地挖掘学业风险学生学习行为模式前后的关联性和周期性,那么模型对正样本特征学习能力不足,最终也就无法捕捉到更多的学业风险学生,进而导致召回率 Recall 相对于禁用多头注意力机制时较低。

4.6 超参数影响

本节主要讨论实验中批量大小(Batch Size)和隐藏维度(Hidden Dimension)对模型性能的影响。为了更好地观察模型的变化趋势,选择准确率 Accuracy 和综合指标 F1 作为衡量模型性能的指标。

1)不同批处理大小对实验结果的影响

在实验中,批量大小的选定范围设定为{32,64,128,256,512,1024},同时确保其他超参数维持其最优配置不变,结果如图 4 所示。

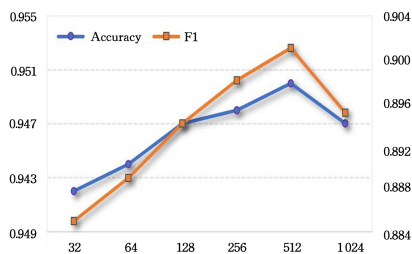


图 4 批量大小的影响

Fig. 4 Influence of batch sizes

从表中可以看出,随着批处理大小的增加,模型性能先上升后下降,当批处理大小为 512 时模型性能最佳。在一定范围内增大批量大小(Batch Size)能够提升模型的预测效能,当 Batch Size 设定为 32 时,模型性能表现不佳,这很可能归因于过小的批量引入了显著的噪声梯度,进而干扰了实验结果的准确性。尽管增大 Batch Size 可以加速训练过程,但相应地,训练周期会延长,且模型性能未必能达到最优状态。如图 4

所示,当 Batch Size 增加至 1024 时,模型性能开始呈现下降趋势。

2)不同隐藏维度对实验结果的影响

实验中选取判别器中隐藏维度的集合为{4,8,16,32,64},同时其他超参数保持最优设置不变。表 3 列出了不同隐藏维度对实验结果的影响。为了精确捕捉隐藏维度变化时模型性能的细微差异,将预测结果的精确度提升至小数点后四位。

表 3 不同隐藏维度大小对实验结果的影响

Table 3 Influence of different hidden dimension sizes on experimental results

隐藏维度	4	8	16	32	64
Accuracy	0.9480	0.9498	0.9498	0.9486	0.9480
F1	0.8976	0.9006	0.9013	0.8982	0.8974

从表 3 中可以看出,当隐藏维度为 4 时,模型表达能力受限,这可能是由于此时无法捕捉到输入数据之间的复杂关系。随着隐藏维度增加至 8 时,模型开始学习到输入数据的主要特征。当增至 16 时,此时模型已经能够对学生的潜在特征进行有效捕捉和充分的学习,进而最终的预测结果达到最佳。当再次逐渐增大时,模型性能开始下降,原因可能是随着隐藏层神经元数量的增多,其复杂度也在增加,导致其过度拟合训练数据中的噪声或异常特征,而忽视了数据的整体分布。

结束语 本研究针对学生学习表现预测领域教育数据不平衡解决方法中存在的问题,提出了一种新的基于生成模型的学习表现预测混合方法 BMVGS。首先,该方法引入基于 BiLSTM 的 VAE 对 GAN 模型进行初始化,使得 GAN 模型能够从一个更加稳定的点开始训练的同时,也能够很好地理解学生学习行为的前后关联性和周期性特征。其次,在 VAE-GAN 模型的判别器中加入多头注意力机制,使得判别器可以更有效地聚焦于输入数据的关键区域,增强其对真实数据和生成数据的区分能力,进而与生成器继续博弈。最后,将基于两者的 VAE-GAN 模型与经典采样方法基于 Blending 集成学习的思想进行融合以提高模型整体的生成能力。在真实学生不平衡数据集上进行了实验,结果表明,所提模型能够生成更高质量的样本,进而提高预测模型对学业风险学生的识别能力。

对于未来的研究工作,可以从两个方面进一步深入研究:

- 1)收集学生利用大模型辅助学习的数据,丰富数据特征以进一步提高模型性能;
- 2)学生表现的早期预测不能仅停留在模型性能的提升上,更应该根据预测的结果给学生提供针对性的解释和提供科学的干预措施,通过动态预测及可解释的干预措施降低学生不及格或辍学比例。

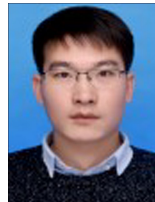
参考文献

- [1] SCHELL J, LUKOFF B, ALVARADO C. Using early warning signs to predict academic risk in interactive, blended teaching environments[J]. *Internet Learning*, 2014, 3(2): 6.
- [2] PATIL P, GANESAN K, KANAVALLI A. Effective deep learning model to predict student grade point averages[C]// *IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*. Coimbatore, India, 2017: 1-6.
- [3] KIM B H, VIZITEI E, GANAPATHI V. GritNet: student performance prediction with deep learning[J]. *arXiv:1804.07405*, 2018.

- [4] YAO L, CUI C R, MA L L, et al. Student performance prediction base on campus online behavior-aware[J]. *Journal of Computer Research and Development*, 2022, 59(8): 1770-1781.
- [5] CHAWLAN V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. *Journal of Artificial Intelligence Research*, 2002, 16: 321-357.
- [6] DU X, YANG J, LI H. An integrated framework based on latent variational autoencoder for providing early warning of at-risk students[J]. *IEEE Access*, 2020, 8(99): 10110-10122.
- [7] SARWAT S, ULLAH N, SADIQ S, et al. Predicting students' academic performance with conditional generative adversarial network and deep SVM[J]. *Sensors*, 2022, 22(13): 4834.
- [8] ZHENG Y F, ZHENG S, DENG M M, et al. MOOC dropout prediction using a fusion deep model based on behaviour features [J]. *Computers and Electrical Engineering*, 2022, 104: 108409.
- [9] ZHANG M Y, DU X, LI H. Research on early warning for learning performance combined with students' behavior patterns analysis [J]. *Computer Engineering and Applications*, 2022, 58(1): 99-105.
- [10] CHEN J, WEI G L, LIU J X, et al. A prediction model of student performance based on self-attention mechanism [J]. *Knowledge and Information Systems*, 2023, 65: 733-758.
- [11] HUANG C, LIU G, JIANG W, et al. Learning Pattern Recognition and Performance Prediction Method Based on Learners' Behavior Evolution [J]. *Computer Science*, 2024, 51(10): 67-78.
- [12] WANG S, NI L, ZHANG Z, et al. Multimodal prediction of student performance: A fusion of signed graph neural networks and large language models [J]. *Pattern Recognition Letters*, 2024, 181: 1-8.
- [13] KINGMA D P, WELING M. Auto-encoding variational bayes [J]. *arXiv*: 1312. 6114, 2013.
- [14] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [J]. *Advances in Neural Information Processing Systems*, 2014, 27(2): 2672-2680.
- [15] RACHBUREE N, PUNLUMJEAK W. Oversampling technique in student performance classification from engineering course [J]. *International Journal of Electrical and Computer Engineering*, 2021, 11(4): 3567-3574.
- [16] RAHIM AA, BUNIYAMIN N. Mitigating imbalanced classification problems in academic performance with resampling methods [J]. *Journal of Electrical and Electronic Systems Research*, 2023, 23: 1985-5389.
- [17] ZHANG Y, LU M. Based on graph-VAE model to predict student's score [J]. *arXiv*: 1903. 03609, 2019.
- [18] CHUI K T, LIU R W, ZHAO M, et al. Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine [J]. *IEEE Access*, 2020, 8: 86745-86752.
- [19] SATHIYAPRIYA S, KANAGARAJ A. Student performance prediction using modified chicken swarm optimization and improved conditional generative adversarial network -with parallel support vector machine over educational data [C] // 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA). Coimbatore, India, 2021: 1-7.
- [20] WAHEED H, ANAS M, HASSAN S U, et al. Balancing sequential data to predict students at-risk using adversarial networks [J]. *Computers and Electrical Engineering*, 2021, 93: 107274.
- [21] LENIN T, CHANDRASEKARAN N. Learning from imbalanced educational data using ensemble machine learning algorithms [J]. *Special Issue on Artificial Intelligence in Cloud Computing*, 2021, 18: 183-195.
- [22] RABELO A M, ZÁRATE L E. A model for predicting dropout of higher education students [J]. *Data Science and Management*, 2025, 8(1): 72-85.
- [23] FAN Z, WANG Y, MENG L, et al. Unsupervised anomaly detection method for bearing based on VAE-GAN and time-series data correlation enhancement [J]. *IEEE Sensors Journal*, 2023, 23(23): 29345-29356.
- [24] JUSTIN L. VRNNGAN: A recurrent VAE-GAN frame work for synthetic time-series [D]. Toronto: Department of Faculty of Information, University of Toronto, 2022.
- [25] NIU Z, YU K, WU X. LSTM-Based VAE-GAN for Time-Series Anomaly Detection [J]. *Sensors*, 2020, 20(13): 3738.



DUAN Chao, born in 1987, Ph.D, lecturer, master supervisor. His main research interests include educational data mining and recommender system.



ZHANG Mingyan, born in 1988, Ph.D, lecturer, master supervisor. His main research interests include educational data and text mining and learning analytics.