

基于GraphSAGE-GRU模型的电子信息专业 学生学习成绩预测研究

孙长智

(亳州学院 电子与信息工程系,安徽 亳州 236800)

摘要:图神经网络算法(graph sample and aggregate, GraphSAGE)和循环神经网络算法(gated recurrent unit, GRU)能利用图结构信息和时间序列信息克服数据稀疏性和高维性的挑战,提高学习行为分析和预测的准确性。通过黑板(Blackboard)平台收集2 000名学生的在线学习时长、作业提交情况和发帖数等学习行为数据,构建了学生学习行为结构,并运用GRU模型对学生的学习历史进行时间序列建模,使用10折交叉验证对模型超参数进行调优。结果表明:学生在不同课程中的及格率存在差异,GraphSAGE-GRU模型在预测学生成绩方面表现突出,均方根误差(RMSE)为0.08,平均绝对误差(MAD)为0.07,决定系数(R^2)达到0.90,明显优于单独使用GraphSAGE或GRU模型;而且学习成绩与学生的在线天数、作业完成率和求助老师的次数呈正相关。研究结果为教学管理提供了数据支持和决策依据。

关键词:GraphSAGE;GRU;学习行为分析;成绩预测;教学决策支持

中图分类号:G642.0 **文献标识码:**A **文章编号:**1671-055X(2026)01-0103-18

在现代教育领域,信息技术的飞速发展和大数据的广泛应用,使得教育数据的分析与预测成为提升教学质量和学生学习效果的重要手段^[1]。电子信息工程专业学生的行为数据种类繁多且复杂,不仅反映了学生的学习态度和习惯,还蕴含着学生在学习过程中的认知水平和学习效果^[2]。然而,现有的学习行为分析与预测方法在应对数据稀疏性和高维性问题时面临挑战^[3],难以捕捉复杂的学习行为特征^[4]。为此,本研究通过Blackboard平台收集2 000名学生的在线学习行为数据,结合图神经网络算法GraphSAGE和循环神经网络

收稿日期:2025-06-12

基金项目:安徽省质量工程项目“FPGA技术”(2024jcs186);安徽省质量工程项目“应用型本科高校电子类课程实验教学数字化资源建设研究”(2023ssyyj053)。

作者简介:孙长智,男,安徽蒙城人,副教授,主要从事电工电子系统设计、通信算法等研究。

络算法 GRU,运用GRU模型对学生的学习历史进行时间序列建模,并使用10折交叉验证对模型超参数进行调优,旨在提出一种基于GraphSAGE-GRU模型的学习行为分析与预测方法,以弥补现有方法的不足,为教学管理提供数据支持和决策依据。

1 数据收集处理与模型介绍

1.1 数据收集和预处理

为了有效分析电子信息工程专业学生的学习行为,数据集从Blackboard平台获取。Blackboard作为广泛使用的学习管理系统,能够详细记录学生在在线学习中的各种行为活动,提供丰富的数据支持^[5]。本次搜集的数据涵盖多个维度,包括在线学习时长、作业提交情况、发帖数、回帖数、测试成绩等。这些数据不仅反映了学生的参与度和学习投入,还提供了其在课程中的实际表现,有助于全面分析学生的学习行为模式。通过查看课程页面上已注册学生名单,单独提取电子信息工程专业的学生数据(为了确保数据的完整性和代表性,数据覆盖从在线学习行为到测试成绩的多个学习活动维度),经过整理与筛选,最终确定2 000名学生的详细数据,其基本数据和学习数据分别如表1和表2所示。

表1 学生的基本数据

年级	性别	人数	比例/%
大一	男	200	10.00
	女	300	15.00
大二	男	280	14.00
	女	260	13.00
大三	女	240	12.00
	男	250	12.50
大四	女	250	12.50
	男	220	11.00

表2 学生的学习数据

收集时间	学生号	学习时间/h	作业状态	发帖数	回帖数	测试结果
2023-08-01	20231001	12.50	已提交	5	8	87
2023-08-01	20231002	8.20	未提交	3	2	85
2023-08-01	20231003	15.00	已提交	7	10	91
2023-08-01	20231004	9.70	已提交	4	3	80
2023-08-01	20231005	13.40	已提交	6	9	82
2023-08-01	20231006	11.20	未提交	2	1	78
2023-08-01	20231007	14.80	已提交	8	11	76
2023-08-01	20231008	7.50	未提交	1	0	56
2023-08-01	20231009	10.30	已提交	5	6	87
2023-08-01	20231010	12.00	已提交	4	5	82

从表1可以看出,学生涵盖了从大一到大四的不同年级,男女比例分布均衡。其中大一年级共有500人——男性200人,女性300人;大二年级男生占比为14.0%,女生占比为13.0%;大三年级男生占比为12.0%,女生占比为12.5%;大四年级则各占12.5%和11.0%。

由于收集的原始数据中可能存在诸多问题,包括数据缺失、异常值和重复数据等,需要进行数据清洗^[6]:对于少量的缺失数据,采用最近邻填充法进行补充,以保留尽可能多的有效信息^[7];而对于缺失严重的记录,则直接删除,避免这些不完整数据影响模型的训练效果。异常值的处理使用四分位距法^[8],识别出那些明显偏离正常范围的异常数据点,并对其进行调整或剔除,确保模型输入的数据合理。对于重复数据,逐一排查并删除重复记录,同时处理可能存在的数据噪声,以提高数据的质量和准确性。最终清洗后的数据能够更准确地反映学生的学习行为类型,如表3所示。

表3 学生学习行为类型

行为类型	操作	数据变量
获取课程行为 自我管理行为 讨论交流行为	登录和注销在线课程	首次访问课程的时间
		在线天数
		在线时间
作业和考试行为 寻求资源行为	添加课程表、查看公告	点击课程表次数
		公告阅读率
获取课程行为	发帖和回复	发帖数
		回帖数
自我管理行为 讨论交流行为	完成作业和参加考试	作业完成率
		测试参与率
作业和考试行为	搜索信息并向教师寻求帮助	搜索信息次数
		向老师寻求帮助的次数

为了进一步提升模型的准确性,对基础特征进行细化和扩展,构建多个衍生特征^[9]。例如,在线学习时长不仅被捕捉为一个整体数据,还进一步细化为每周和每日的学习时长,以便更好地反映学生的学习习惯和时间分布规律。类似发帖数和回帖数结合具体的课程内容进行深入分析,能够提取出学生在特定主题下的参与情况。这种特征的细化和衍生有助于更全面和深入地理解学生的学习行为,揭示学生的学习态度和习惯背后的潜在规律。

在此基础上,整合来自不同数据源的数据,形成完整的学习行为特征矩阵^[10]。为了确保模型在处理不同量纲的数据时不受某一特征的过度影响,采用了特征归一化和标准化的方法。通过这些处理步骤,不同维度的特征数据被转化为统一尺度,从而避免了因量纲差异导致的模型不均衡现象。处理后的数据被输入到 GraphSAGE 和 GRU 模型中进行训练与预测。GraphSAGE 模型负责捕捉学生之间的互动和图结构信息,而 GRU 模型则利用时间序列信息进行时序预测,这种组合模型能够有效提升学习行为分析与成绩预测的准确性。

1.2 学习行为图表示

在电子信息工程专业学生的学习行为分析与预测中,图结构的构建不仅能够直观地表示学生之间的互动关系,而且能够为 GraphSAGE 模型提供必要的输入数据^[11]。图结构的构建过程包括节点定义、边的生成与权重设定,以及图结构的优化与存储。每个学生被定义为图中的一个节点,每个节点包含学生的标识信息以及一系列与其学习行为相关的特征数据^[12]。边的生成主要基于学生在 Blackboard 平台上的互动行为——如果两个学生共同参与了同一主题的讨论,则这两个学生之间的互动行为被视为一条边。连接其各自对应的节点,边的权重表示学生之间的互动频率,反映学生关系网络的动态变化。学生学习行为图的构建结果如图1所示。

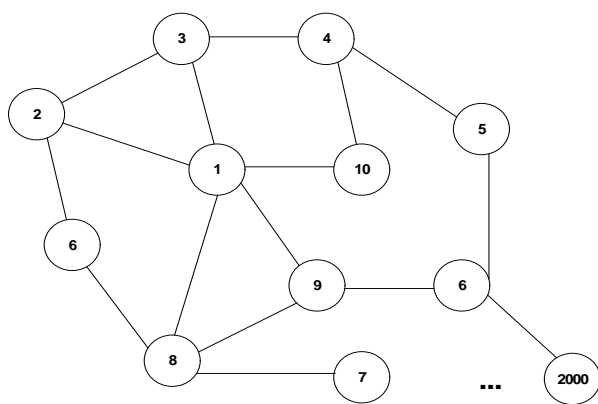


图1 学生学习行为图结构

GraphSAGE 模型通过采样和聚合节点的邻居特征,逐层生成节点的表示向量,采用节点自身的特征信息同时综合其邻居节点的特征信息,捕捉到图结构中的局部和全局信息^[13]。其采样和聚合过程如图2所示。

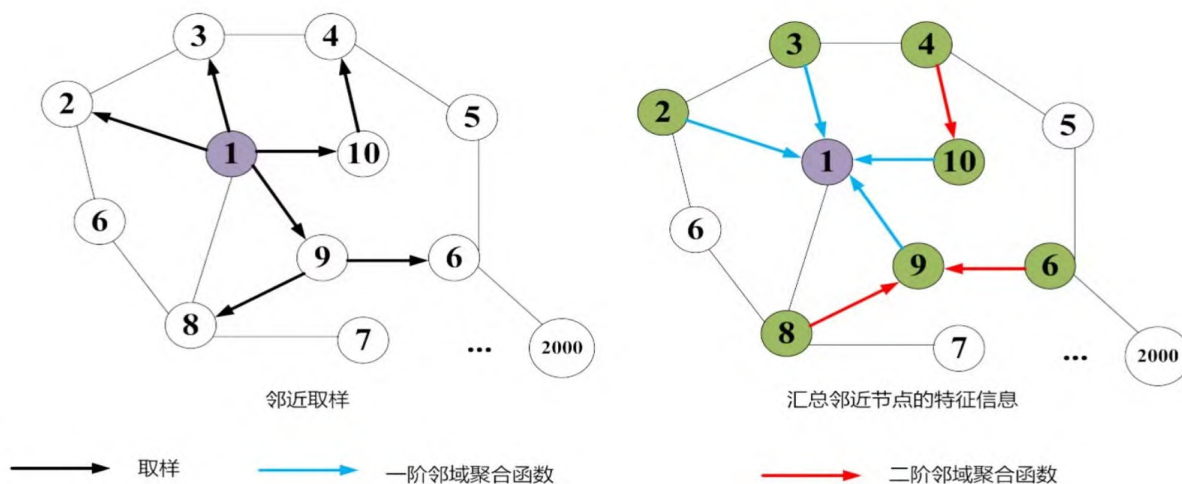


图2 GraphSAGE 采样和聚合过程

GraphSAGE 模型^[14]通过分层神经网络结构逐步学习每个节点的嵌入表示。在第 k 层,节点 v 的表示为 $h_v^{(k)}$,对于每个节点 v ,从其邻居节点集合 $N(v)$ 中采样一部分邻居节点,记作 $N_s(v)$ 。

采样完成后,GraphSAGE 模型通过聚合函数对这些邻居节点的特征进行聚合。第 k 层中节点 v 的邻居特征聚合表示为式(1):

$$h_{N(v)}^{(k)} = \text{Aggregate}\{[h_u^{(k-1)}, \forall u \in N_s(v)]\} \quad (1)$$

邻居节点 $u \in N_s(v)$ 的特征表示为 $h_u^{(k-1)}$ 。

在得到聚合的邻居节点特征后,GraphSAGE 结合节点 v 自身的特征和邻居节点的聚合特征,更新节点 v 的表示向量。更新操作通过一个非线性变换来实现,表示为式(2):

$$h_v^{(k)} = \sigma[W^{(k)} \cdot \text{Concat}(h_v^{(k-1)}, h_{N(v)}^{(k)})] \quad (2)$$

式中: $W^{(k)}$ 表示第 k 层的权重矩阵; Concat 表示将节点自身的特征和邻居特征进行拼接。

经过多层迭代,每个节点 v 的表示向量 $h_v^{(k)}$ 是在第 k 层得到的嵌入表示。嵌入表示不仅包含了节点 v 自身的学习行为特征,还融入了其邻居节点的行为特征信息,提供了对图结构的全局视角和上下文信息。

1.3 时序特征建模

在生成学生图嵌入表示向量后,将 GraphSAGE 生成的每个学生的嵌入表示向量与该学生的历史学习行为数据结合。这些历史数据包括多维特征,如在线学习时长、作业提交情况、发帖数、回帖数等,形成一个综合特征矩阵。该矩阵既反映了学生在图结构中的关系信息,又包含了学生在多个时间步长中的具体学习行为表现。

GRU 模型^[15]是一种循环神经网络的变体,专门设计用于处理和预测时间序列数据。GRU 通过引入更新门和重置门,解决了在处理长序列数据时可能遇到的梯度消失或梯度爆炸问题^[16]。在每个时间步长 t ,GRU 接收输入向量 x_t 和前一时刻的隐藏状态 h_{t-1} ,更新的公式为式(3):

$$z_i = \sigma[W_z \cdot (h_{i-1}, x_i) + b_z] \tag{3}$$

式中： W_z 表示权重矩阵； b_z 是偏置项； σ 是 Sigmoid 激活函数。更新门控制了前一时刻的信息如何在当前时刻进行更新。

重置门表示为式(4)：

$$r_i = \sigma[W_r \cdot (h_{i-1}, x_i) + b_r] \tag{4}$$

重置门决定了前一时刻的隐藏状态 h_{i-1} 中有多少信息需要被遗忘。

候选隐藏状态包含了当前时刻对隐藏状态的更新信息。候选隐藏状态表示为式(5)：

$$\tilde{h}_i = \tan h [W_h \cdot (r_i * h_{i-1}, x_i) + b_h] \tag{5}$$

隐藏状态更新表示为式(6)：

$$h_i = (1 - z_i) * h_{i-1} + z_i * \tilde{h}_i \tag{6}$$

最终的隐藏状态 h_i 是由前一时刻的隐藏状态 h_{i-1} 和候选隐藏状态 \tilde{h}_i 加权组合而成的，其中更新门 z_i 决定了二者的加权比例。

GRU 在每一个时间步长上生成一个新的隐藏状态 h_i ，该状态综合了学生的当前学习行为特征和过去的行为特征^[17]，不仅保留了时序数据的动态变化信息，还有效地应对了长时间跨度内的依赖关系，从而在时序数据建模中展现出强大的表现力。在 GRU 模型中，学生的历史学习行为数据被组织成一个时间序列，这些时间序列数据通过 GRU 的多个循环层逐步提取时序特征。每一层 GRU 都会在接收上层输出的基础上，进一步提取更深层次的时间依赖关系信息。GRU 模型输出一个预测值，表示对学生未来学习成绩的预测。初始设置的 GRU 模型参数如表 4 所示。

表 4 初始设置的 GRU 模型参数

参数	值	参数	值
Input dim	128	Dropout ratio	0.500
Hidden dim	256	Learning rate	0.001
Output dim	256	Optimizer	Adam
Number of layers	2	Batch size	64.000

1.4 GraphSAGE-GRU 模型

将 GraphSAGE 生成的节点表示与学生的历史学习行为数据结合，然后将其输入到 GRU 模型中进行时序预测^[18]，构建的 GraphSAGE-GRU 模型如图 3 所示。

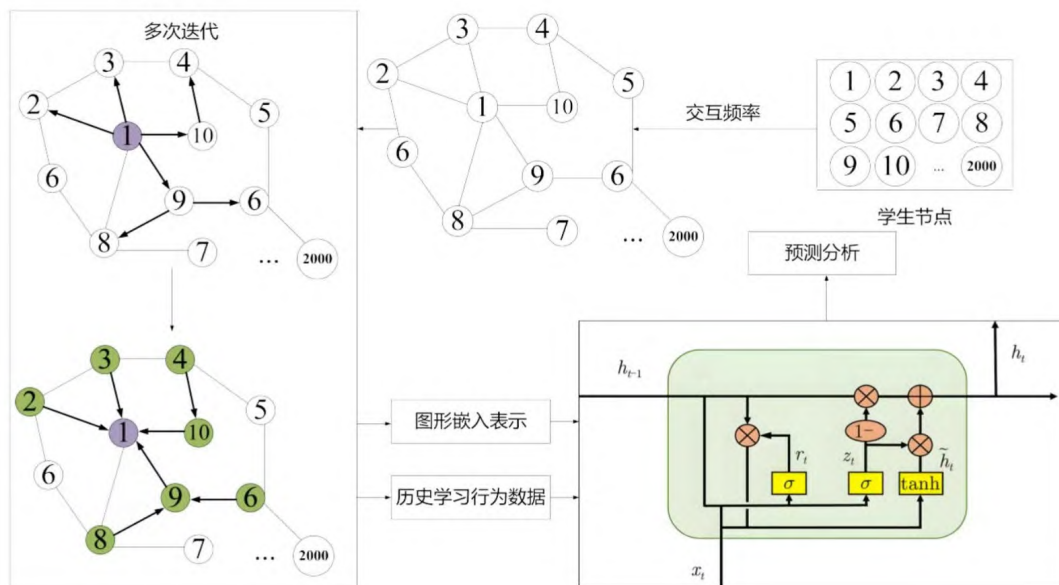


图3 GraphSAGE-GRU 模型

构建的 GraphSAGE-GRU 模型旨在对电子信息工程专业学生的学习行为进行深入分析与精准预测^[19]。通过将 GraphSAGE 生成的节点嵌入表示与 GRU 模型结合,形成了一个端到端的学习行为分析与预测模型。该模型能够充分利用图结构信息与时序特征,有效提升预测的准确性。为了确保模型的性能,数据集按照 3:1:1 的比例划分为训练集、验证集和测试集。在验证集中,采用 10 折交叉验证对模型进行调优——主要针对 Dropout ratio 和 Learning rate 等关键参数进行调整和优化,调优后的参数结果如表 5 所示。

表 5 模型参数调优的结果

Dropout ratio(随机失活率)	Learning rate(学习率)	Loss value(损失值)
0.1	0.010 0	0.023 0
0.1	0.001 0	0.028 0
0.1	0.000 1	0.035 0
0.2	0.010 0	0.019 0
0.2	0.001 0	0.024 0
0.2	0.000 1	0.032 0
0.3	0.010 0	0.017 0
0.3	0.001 0	0.021 0
0.3	0.000 1	0.028 0
0.4	0.010 0	0.016 0
0.4	0.001 0	0.020 0
0.4	0.000 1	0.027 0
0.5	0.010 0	0.015 0
0.5	0.001 0	0.019 0
0.5	0.000 1	0.026 0

通过这种方法,模型在处理稀疏性和高维数据时具备了更强的鲁棒性与预测能力,确保了模型在不同数据集上的稳健性和泛化能力。

损失值通过均方误差计算获取,当Dropout ratio为0.5且Learning rate为0.010 0时损失值达到最低0.015 0。均方误差(Mean Square Error, MSE)的公式为式(7):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (7)$$

式中: \hat{y}_i 和 y_i 分别表示模型的预测值和实际值; n 为样本数量。

在训练过程中,通过反向传播和随机梯度下降算法,迭代更新模型的参数,以最小化损失函数^[20]。模型训练完成后,使用测试集对模型的预测性能进行评估^[21]。均方根误差(Root Mean Square Error, RMSE)的公式为式(8):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (8)$$

平均绝对误差(Mean Absolute Deviation, MAD)的公式为式(9):

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (9)$$

决定系数(R^2)衡量模型式(10)解释方差的能力,值域为0~1,接近1时表示模型的拟合效果好。

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

2 结果与分析

2.1 学习行为分析

在某“211”高校电子信息工程专业选取10门课程,其中必修专业课5门,分别为“信号与系统”“数字电子技术”“通信原理”“电磁场与波”“自动控制原理”;选修专业课5门,分别为“嵌入式系统设计”“数字信号处理”“射频与微波技术”“集成电路设计”“物联网技术与应用”。

通过收集的数据,分析得出专业必修课的及格率,统计结果如图4所示。

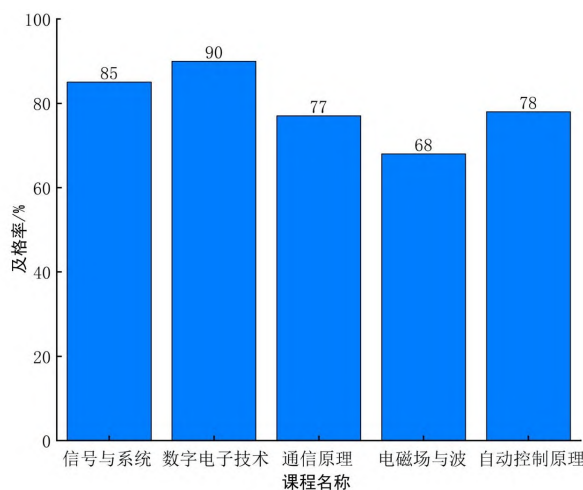


图4 电子信息工程专业必修课程及格率统计

及格率的差异反映了课程对学生的挑战程度,学生对课程知识与技能的掌握情况。在电子信息工程专业的必修课中,“信号与系统”“数字电子技术”两门课程的及格率超过80%,”通信原理”“自动控制原理”两门课程及格率接近80%,而“电磁场与波”课程的及格率低于70%。此外,该专业选修课程的人数及其及格率如图5所示。

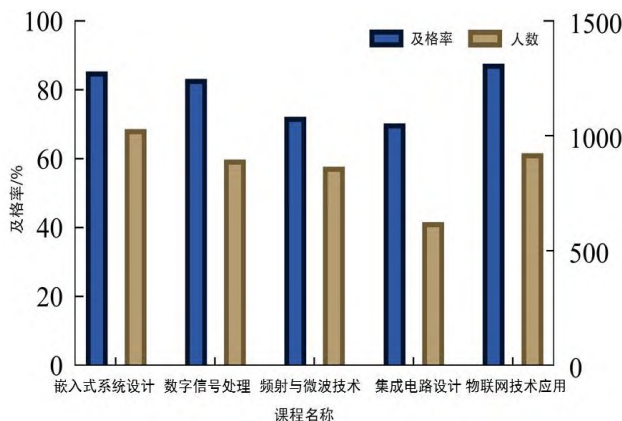


图5 电子信息工程专业选修课程人数及其及格率

在图5中,及格率和选修人数的分布显示了学生对不同领域的兴趣与挑战程度。“物联网技术及应用”课程以87.60%的及格率居于前列,表明该领域的实用性和学生的兴趣高;“嵌入式系统设计”和“数字信号处理”课程也分别具有85.40%和83.20%的及格率,吸引了1029人和897人选修,反映了这些课程的广泛应用性和相对较高的及格率;“射频与微波技术”和“集成电路设计”课程的及格率较低,分别为72.20%和70.30%,选修人数也相对较少,分别为865人和624人,这可能是因为这些课程涉及的技术内容较为复杂。上述数据显示,虽然及格率高的选修课程通常较为受欢迎,但对于技术难度较大的课程,学生可能需要更多的学习支持,各门课程的及格率整体分布在70%~90%之间,差异较为明显。同时,该专业选课人数较多的课程,及格率大多在80%以上,这说明学生在选课前了解了有关课程的难易度,倾向于选择通过率高的课程。电子信息工程专业学生在线天数对应的人数如图6所示。

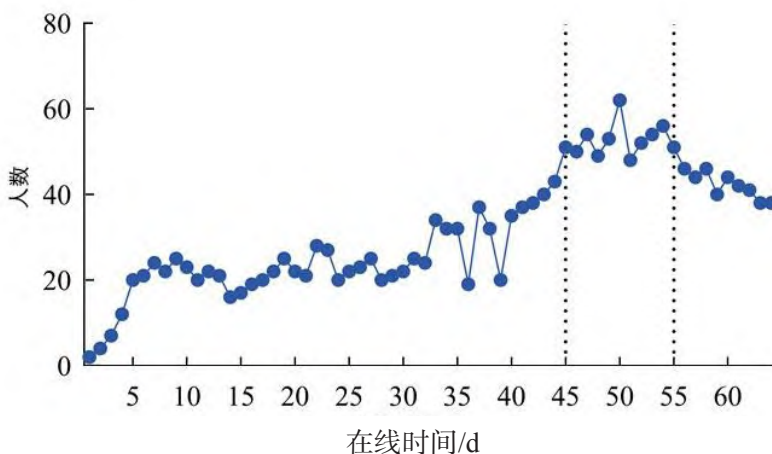


图6 电子信息工程专业学生在线时间分布

图6显示了不同在线天数对应的人数分布情况:随着在线天数的增长,人数大体上呈现先增长后下降的趋势。在线天数为45~55范围内的人数相对较多,这一区间的人数均超过了50人,并在在线天数为50时达到最大值62人。学生发帖数、回帖数的结果如表6所示。

表6 电子信息工程专业学生课程在线学习发帖数、回帖数结果

发帖数	人数	回帖数	人数
0	240	0	430
1~5	580	1~5	510
6~20	500	6~20	480
21~50	320	21~50	350
51~100	200	51~100	166
100及以上	160	100及以上	64

发帖数和回帖数的6个区间,反映了学生的活跃程度和参与度。从表6可以看出,大部分学生的发帖数和回帖数都集中在较低区间,说明大多数学生的参与度较为有限。发帖数为0的学生有240人,而回帖数为0的学生更多,达到430人,这表明有相当一部分学生在课程论坛中几乎没有互动。虽然发帖和回帖数量在1~5、6~20之间的学生人数相对较多,但随着互动频率的增加,人数明显减少,这表明少数学生更加活跃,发帖和回帖频繁。学生作业完成质量与对应的成绩如表7所示。

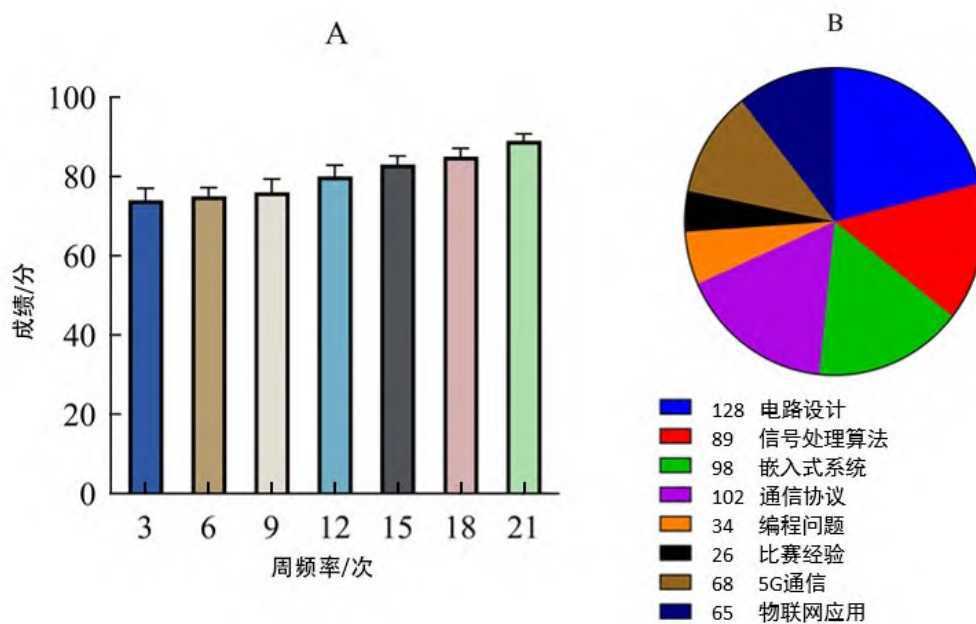
表7 电子信息工程专业学生作业完成质量与对应的成绩

完成质量	人数	平均成绩	方差
优秀	250	91.50	2.10
良好	809	84.30	2.70
中等	691	75.60	2.80
差	250	56.70	3.20

从表7可以看出,电子信息工程专业学生作业完成质量与平均成绩呈现明显的正相关关系:作业完成质量为“优”的学生,人数为250人,平均成绩达到91.50分,且方差较小,表明这些学生的成绩相对稳定且表现出色;作业质量为“良”和“中”的学生分别有809人和691人,平均成绩为84.30分和75.60分,方差略微增加,说明成绩有所波动但仍处于较高水平;作业质量为“差”的学生有250人,平均成绩仅为56.70分,且方差最大,反映这些学生的成绩不仅较低且不稳定。总体来看,作业完成质量越高,学生的平均成绩越好,且成绩的稳定性也越好。

2.2 学习的互动关系分析

通过 GraphSAGE 捕捉到学生之间的关系强度,通过互动频率来体现关系强度。电子信息工程专业不同互动频率所对应的成绩信息以及互动内容信息如图 7 所示。



注:A为不同互动频率所对应的成绩信息,B为互动内容信息。

图7 学生互动频率和内容分布

从图 7 可以看出,随着每周互动次数的增加,学生的平均成绩逐步提升。例如,每周互动 3 次的学生平均成绩为 74 分,而每周互动 21 次的学生平均成绩则达到了 89 分。这一趋势表明,频繁的互动不仅能够帮助学生更好地理解课程内容,还能够通过与同学或教师的讨论,及时解决学习中的疑问,进而有效提高学习效率和成绩。频繁的互动能够促进学生对知识的深度理解,能够帮助学生在知识点上构建更加扎实的体系。因此,鼓励学生增加互动,尤其是在学习难度较高的科目中显得尤为重要,这一策略有助于提升整体的学习效果。

在一周内的互动内容分布上,不同主题的互动次数反映了学生对各个学习领域的关注点与需求差异:电路设计的互动次数最高,达 128 次,表明该课程由于概念复杂且实践性强,学生在学习过程中需要更多地讨论与交流;“信号处理算法”“嵌入式系统”“通信协议”的互动次数也相对较高,分别为 89、98 和 102 次,表明这些技术性较强且涉及多个知识点的课程同样需要频繁的探讨;而对于“编程问题”“比赛经验”“5G 通信和物联网应用”等课程内容的互动次数相对较少,分别为 34、26、68 和 65 次,表明这些领域对大部分学生而言相对容易掌握,或者是在这些领域中遇到的疑问较少。总体来看,学生在复杂且理论性强的科目中更倾向于增加互动,这也强调了这些课程在学习过程中所需深度理解与广泛交流的重要性。

2.3 学习时间分布

分析不同年级学生每天的学习时长及学习高峰时间段,可以全面了解学生的学习习惯和行为模式,从而制定有效的教学策略和支持措施。学生每天学习时长及学习高峰时间段如表8所示。

表8 电子信息工程专业学生每天学习时长及学习高峰时间段

年级	平均日学习时间/h	学习高峰期
大一	5.60	9:00~12:00
大二	6.80	13:00~16:00
大三	6.60	13:00~16:00
大四	5.40	14:00~16:00

从表8可以看出,大二和大三的学生学习时长相对较长,表明该阶段学生学习压力较大,面临核心课程和专业课程的学习任务。相比之下,大一和大四的学生学习时长相对较短。在学习高峰时间段方面,大一学生的高峰时间段为上午9:00~12:00,这一时间段的选择与大一新生的课表安排较为集中在上午有关;大二和大三学生的学习高峰时间段都集中在下午13:00~16:00,这与完成下午的实验课程或课后作业有关;大四学生的学习高峰时间段则推迟到了14:00~16:00,反映了大四学生学习节奏的放缓和更为自主的学习安排。

不同年级学生学习时长和学习高峰时间段的差异,反映了学生在大学不同阶段的学习需求和行为模式。对于大一新生,学校可以在上午安排更多基础课程,以帮助学生养成良好的学习习惯;而对于大二和大三的学生,学校可以加强下午时段的教学资源和支持,通过增加辅导课或自习时间,以满足高效学习需求;对于大四学生来说,学校可以提供更多的自主学习和实践机会,让学生顺利过渡到毕业后的职业生涯。

2.4 学习成绩分析

GraphSAGE捕捉学生之间复杂的非线性关系,GRU捕捉学生成绩变化的动态特征。将GraphSAGE-GRU结合,把社交网络中的静态特征与时间序列中的动态特征整合,全面预测学生学习成绩。

首先收集各个课程或群体的学生成绩数据,并把这些数据分为不同的组别。接着对每组数据计算平均成绩,并进一步应用方差公式来衡量成绩分布的离散程度,即计算每个学生成绩与该组平均成绩之差的平方和的平均值。通过比较不同组别的方差,能够了解各组成绩的波动情况,识别出哪些课程或学生群体的成绩更加一致或分散。最后直观地展示不同课程或学生群体之间成绩分布的差异。这种分析对于教育管理者来说非常有价值,有助于识别教学中可能存在的问题,并为教学改进和资源分配提供数据支持。学习成绩预测结果如图8所示。

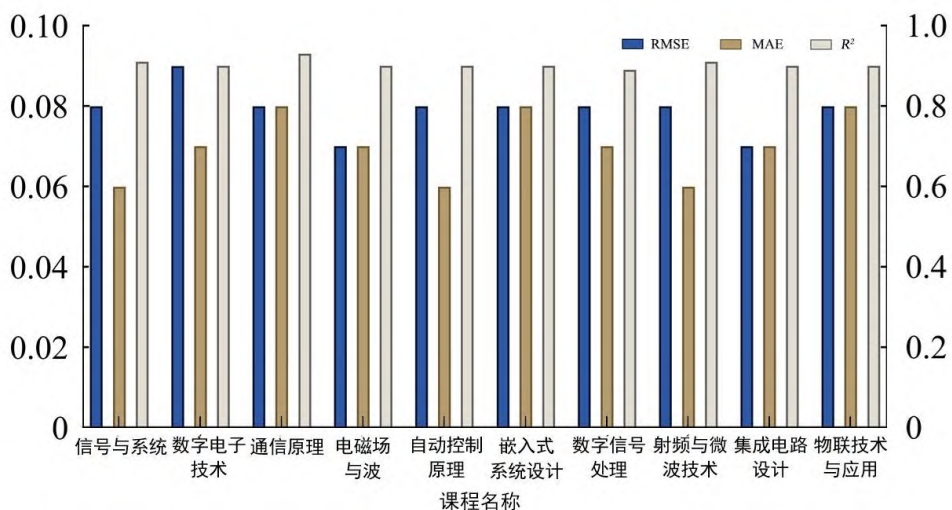


图 8 学生学业课程名称成绩预测结果

从图 8 可以看出,“通信原理”课程的 R^2 值达到 0.93, 表明模型能够解释 93% 的成绩变化, 预测性能优异。这门课程的 MAD 和 RMSE 值均为 0.08, 提示预测过程中仍有部分误差, 这是课程内容的复杂性和学生掌握程度的差异性所致。“电磁场与波和集成电路设计”课程的预测表现较为出色, RMSE 和 MAD 均为 0.07, R^2 值也达到了 0.90, 表明模型在这些课程中能精确捕捉成绩的变化模式。“数字电子技术”“自动控制原理”“嵌入式系统设计”等课程的 R^2 值均为 0.90, RMSE 和 MAD 在 0.06 到 0.09 之间波动, 显示出较为稳定的预测精度, 预测一致性与这些课程在结构上较为系统, 以及学生的学习行为和成绩较为规律有关, 使得模型能够较为准确地预测成绩波动。GraphSAGE 结合 GRU 模型在这些课程中的预测性能较好, 且各个指标之间的差异较小, 反映了模型的稳健性。将 GraphSAGE-GRU 与其他模型进行比较, 结果如表 9 所示。

表 9 GraphSAGE-GRU 与其他模型比较结果

模型	RMSE	MAD	R^2
GraphSAGE-GRU	0.08	0.07	0.90
GraphSAGE	0.12	0.10	0.85
GRU	0.09	0.08	0.88
LSTM	0.09	0.09	0.87

从表9可以看出,GraphSAGE-GRU模型在学生学习成绩预测中表现出显著的优势:GraphSAGE-GRU模型的RMSE为0.08,MAD为0.07, R^2 达到0.90;GraphSAGE模型的RMSE为0.12,MAD为0.10, R^2 为0.85,整体预测效果较GraphSAGE-GRU稍差,这可能是由于GraphSAGE在捕捉时间序列信息方面存在不足;GRU模型的RMSE为0.09,MAD为0.08, R^2 为0.88,虽然预测性能优于单独的GraphSAGE模型,但在与GraphSAGE-GRU模型相比时略显逊色,表明GRU在捕捉复杂数据特征时效果略差;LSTM模型的RMSE和MAD均为0.09, R^2 为0.87,显示出较为良好的预测效果,但与GraphSAGE-GRU相比,LSTM的表现稍显不足,这可能因为LSTM模型对长时间依赖性的处理能力较强,但在结合图结构信息时表现不如GraphSAGE-GRU。总体来说,GraphSAGE-GRU模型通过结合图结构学习能力和时间序列分析能力,在处理复杂的学生成绩预测任务中表现出了更强的优势。学习成绩区间分析结果如表10所示。

表10 电子信息工程专业学习成绩区间分析结果

分数区间	平均在线时间/d	平均作业完成率/%	教师平均求助次数
0~60	29	62.20	2
61~70	38	68.90	5
71~80	42	83.40	9
81~90	47	88.80	11
91~100	55	92.40	12

从表10可以看出,学生的在线天数、作业完成率和求助老师的次数与成绩呈现出正相关关系:成绩在0~60分区间的学生,平均在线时间为29 d,平均作业完成率为62.20%,求助老师的次数较少,平均仅为2次,反映出这些学生的学习参与度和自我驱动性较低,导致成绩偏低。随着成绩区间的提高,学生的平均在线天数和平均作业完成率逐渐增加,在61~70分区间,平均在线时间上升至38 d,平均作业完成率提升至68.90%,平均求助次数增至5次,说明这些学生在学习上的投入有明显增加。尤其是在71~80分和81~90分区间,平均在线时间分别达到42 d和47 d,平均作业完成率分别提升至83.40%和88.80%,平均求助次数也相应增加至9次和11次,这表明随着学习投入和教师互动的增加,学生的成绩显著提高。成绩在91~100分区间的学生表现最为突出,平均在线时间55 d,平均作业完成

率高达 92.40%，平均求助次数达到 12 次，显示出对学习的高度重视以及积极的学习行为。总体来说，在线天数、作业完成率和求助次数的增加对学生学习成绩的提升具有积极影响，建议进一步鼓励学生增加在线学习时间和与教师的互动，以提高学习效果。

3 研究结论与建议

3.1 研究结论

通过构建基于 GraphSAGE-GRU 模型的学习成绩预测框架，深入分析了电子信息工程专业学生学习行为与学业成绩之间的关联。研究发现，学生在线学习时长、作业完成率以及与教师的互动频率是影响学业成绩的关键因素。GraphSAGE-GRU 模型在预测学生成绩方面表现出色，其 RMSE 为 0.08，MAD 为 0.07， R^2 达到 0.90，显著优于单独使用 GraphSAGE 或 GRU 模型。这一结果证实了结合图神经网络和循环神经网络在处理教育数据时的有效性，尤其是在捕捉学生间互动关系和时序行为特征方面。此外，研究还发现，学生不同课程中的及格率存在显著差异，这可能与课程难度、学生兴趣和教学方法有关。这些发现为教育管理者提供了数据支持，有助于他们识别教学中的问题，优化教学资源分配，并制定更有效的教学策略。

3.2 建议

通过结合 GraphSAGE 和 GRU 模型，对电子信息工程专业学生的学习行为进行了全面的分析和预测。GraphSAGE 模型通过图神经网络捕捉学生之间的互动关系，聚合邻居节点的信息生成嵌入表示，从而揭示图结构中的复杂关系；GRU 模型处理时间序列数据，精准捕捉学习行为的时序特征，有效提升了对未来学习成绩的预测准确性。进一步的分析表明，学习成绩与在线学习天数、作业完成率以及与教师互动次数存在显著的正相关关系：在线学习天数越多、作业完成率越高以及与教师互动频率越高的学生，成绩表现越好。

3.2.1 促进学生互动与合作学习

高频互动与学生成绩存在高方差相关联，建议教师设计更多促进学生间互动的课程活动，如小组讨论、同伴学习和项目合作。这些活动不仅能提高学生的参与度，还能通过集体智慧解决学习中的难题，从而提升学习效果。此外，教师还可以利用智能教育平台提供的数据分析，及时调整互动形式和内容，以满足学生的不同需求。

3.2.2 优化学习时长与学习效率

不同年级学生学习时长与成绩存在一个最优学习时长。建议学校和教师引导学生合理安排学习时间，避免因过度学习导致效率下降。可以通过工作坊和辅导课程教授学生时间管理技巧，帮助学生找到适合自己的学习节奏，以提高学习效率；同时，利用教育技术工具监测学生的学习时长和效率，及时给予反馈和建议。

3.2.3 调整课程难度与教学方法

对于难度较高的课程,建议教师采用分层次教学,将复杂概念分解为易于理解的小部分,并辅以实际案例和实践操作。此外,可以利用图形化和互动式教学工具,帮助学生更好地理解抽象概念,提高学习兴趣和成绩。教师还应定期收集学生反馈,灵活调整教学内容和方法,以确保课程难度与学生能力相匹配。

3.2.4 提升模型预测性能与教学反馈

GraphSAGE-GRU模型的预测误差分析,可以识别模型存在的不足,并针对性地优化模型;同时,可以将模型预测结果与实际教学反馈相结合,为教师提供及时的学生学习进展信息,帮助调整教学内容和方法,以更好地满足学生的学习需求。此外,可以建立一个教育决策支持系统,利用大数据分析为教育管理者提供科学的决策依据,提升教育管理的效率和效果。

3.2.5 实施个性化教学与优化资源分配

建议学校为后进学生提供个性化辅导,如组织一对一辅导或小组辅导,以解决学生的学习困难。同时,为高互动学生群体提供更多的讨论和合作学习机会,以促进深入学习和创新思维。此外,建议合理分配教学资源,为需求较大的课程提供更多的教学时间和设施支持,并结合人工智能技术,开发适合学生使用的个性化学习平台,以满足不同学生的学习需求。

总之,GraphSAGE-GRU模型将图神经网络与循环神经网络相结合,用于预测电子信息工程专业学生的学业成绩。通过对2 000名学生的多维度学习行为数据进行深入分析,不仅阐明了学生成绩与在线学习天数、作业完成率以及师生互动频率之间的显著相关性,而且通过模型的精确预测为教育实践提供了坚实的数据基础。同时,研究凸显了在教学设计中重视学生互动的重要性,特别是在涉及复杂概念和技术的课程中设计多样的互动活动,有助于提升学习成绩。此外,针对不同年级学生学习习惯的差异性,需要设计个性化的教学策略。

4 参考文献

- [1] 常建华,张秀再. 基于OBE理念的实践教学体系构建与实践:以电子信息工程专业为例[J]. 中国大学教学,2021(Z1):87-92.
- [2] 宋晓华. 民办高校电子信息工程专业应用型人才培养途径[J]. 黑龙江科学,2020,11(19):64-65.
- [3] 牟智佳. 多模态学习分析:学习分析研究新生长点[J]. 电化教育研究,2020,41(5):27-32.
- [4] 张进良,魏立鹏,刘斌. 智能化环境中基于学习分析的学习行为优化研究[J]. 远程教育杂志,2020,38(2):69-79.
- [5] Liang Xinyi, Xing Hongyan, Gu Wei, et al. Hybrid gaussian network intrusion detection method based on CGAN and E-GraphSAGE[J]. Instrumentation, 2024, 11(2): 24-35.

- [6] 孙希明,夏卫国,王伟,等. 提高电子信息类专业研究生教育国际化质量和水平的探索与实践[J]. 工业和信息化教育,2024(8):11-14.
- [7] 李杰,程鸿芳,张磊. 基于 Stacking 集成学习模型的学生学习成绩预测研究[J]. 科技风,2024(26):33-35.
- [8] 黄春利,刘桂梅,姜文君,等. 基于行为演化的学习模式识别及效果预测方法[J]. 计算机科学,2024,51(10):67-78.
- [9] Eka M, Mediana A, Mialka R, et al. Machine learning's model-agnostic interpretability on the prediction of students' academic performance in video-conference-assisted online learning during the covid-19 pandemic[J]. Computers and Education: Artificial Intelligence, 2024, 7: 100312.
- [10] Ashima K, Rajni M, Aman S, et al. A novel methodology using RNN + LSTM + ML for predicting student's academic performance [J]. Education and Information Technologies, 2024, 29 (11) : 14365-14401.
- [11] 李菲,曹阳,顾问. 学生在线学习行为分类融合方法的研究与应用[J]. 电脑与电信,2023,1(8):27-31.
- [12] 敬容,杨逸民,万福成,等. 基于 GraphSAGE 网络的藏文短文本分类研究[J]. 中文信息学报,2024,38(9):58-65.
- [13] 刘啸宇,廖志芳,谈遂,等. 基于堆叠 GRU 神经网络的桥梁动应变预测[J]. 计算机工程,2026,52(3):441-450.
- [14] 牟智佳. MOOCs 环境下学习行为群体特征分析与学习结果预测研究[J]. 中国医学教育技术,2020,34(1):1-6.
- [15] Krishnan S, Jothi J, Arul A ,et al. Automatic disease prediction from human gut metagenomic data using boosting GraphSAGE[J]. BMC Bioinformatics,2023,24(1):126.
- [16] Zhang Wengang, Li Hongrui, Tang Libin, et al. Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks[J]. Acta Geotechnica,2022,17(4):1367-1382.
- [17] 白琳,陈彦萍,潘晓英. 电子信息类专业学位研究生“1234”工程实践创新能力培养模式[J]. 计算机教育,2024(8):14-18.
- [18] 侯艳,李宗睿,杨晓冬. 应用型本科高校电子信息类专业人才培养模式改革探索与实践[J]. 中国管理信息化,2024,27(12):218-220.
- [19] 王硕,杨昱,黄琼,等. 基于融合注意力机制的并列 GRU 应用层协议识别方法[J]. 上海电力大学学报,2024,40(4):377-382.
- [20] 傅常浩,侯秀丽,周君华. 基于决策导向评价模型的体育大单元教学质量评价指标体系构建[J]. 六盘水师范学院学报,2024,36(4):109-120.
- [21] 张敬奎,张佳凯,范毅,等. 案例教学法在工科院校专业课程教学中的应用研究[J]. 教育理论与实践,2024,44(36):57-60.

Research on Predicting Academic Performance of Electronic Information Majors Based on GraphSAGE-GRU Model

Sun Changzhi

(School of Electronics and Information Engineering, Bozhou University, Bozhou 236800, China)

Abstract: Graph Neural Network algorithms (Graph Sample and Aggregate, GraphSAGE) and Recurrent Neural Network algorithms (Gated Recurrent Unit, GRU) can leverage graph structural information and time-series information to overcome challenges like data sparsity and high dimensionality, thereby improving the accuracy of learning behavior analysis and prediction. Learning behavior data (including online learning duration, homework submission status, and forum post counts) from 2 000 electronic information majors was collected via the Blackboard platform. A student learning behavior structure was constructed. The GRU model was applied to model the students' learning history in a time-series manner, and 10-fold cross-validation was used to optimize the model hyperparameters. The results revealed three key findings: (1) Student pass rates exhibited significant variation across different courses; (2) the GraphSAGE-GRU model demonstrated superior predictive performance, with an RMSE of 0.08, MAD of 0.07, and R^2 of 0.90, significantly better than GraphSAGE or GRU models alone; (3) academic performance was positively correlated with the number of online days, homework completion rate, and the number of times seeking help from teachers. These findings provide data support and decision-making references for optimizing teaching management in electronics and information technology programs in higher education institutions.

Keywords: GraphSAGE, GRU; Learning behaviour analysis; Performance prediction; Teaching decision-making support

[责任编辑:江 伟 杨 洪]