

基于改进 Transformer 的持续血糖浓度预测模型

徐鹤^{1,2}, 杨丹丹¹, 刘思行³, 季一木^{1,2}

(1. 南京邮电大学计算机学院, 南京 210023; 2. 江苏省高性能计算与智能处理工程研究中心, 南京 210023; 3. 江苏鱼跃医疗设备股份有限公司, 镇江 212300)

摘要: 糖尿病是一种普遍存在的慢性疾病, 做好血糖控制对糖尿病的预防具有重要作用。然而, 持续血糖监测 (Continuous glucose monitoring, CGM) 过程中数据的不确定性显著增加了血糖预测的难度。因此, 提出一种新的基于深度学习的血糖浓度预测模型, 旨在提高模型对传感器提取数据的适应性。在该模型中, 堆叠式降噪自编码器 (Stacked denoising auto encoder, SDAE) 被嵌入 Transformer 编码器的结构中, 实现对输入数据的重构去噪和特征提取; 然后, 采用混合位置编码策略替代原来的单一绝对位置编码嵌入, 同时将轻量级解码器引入 Transformer 模型中, 替代原始结构复杂的解码器, 聚合来自不同层次的特征信息, 同时获取局部和全局特征; 最后, 通过搭建的 SDAE-改进 Transformer 网络对 CGM 数据序列并行化训练, 更全面地捕捉数据中的时序模式和复杂关联, 提高预测性能。实验结果表明, 该模型相较于传统方法在血糖预测任务中取得了显著的性能提升, 证实了其在处理 CGM 数据时的有效性和鲁棒性。

关键词: 持续血糖监测; 神经网络; 堆叠降噪自编码器; Transformer; 注意力机制

中图分类号: TP18 **文献标志码:** A

Continuous Blood Glucose Concentration Prediction Model Based on Improved Transformer

XU He^{1,2}, YANG Dandan¹, LIU Sixing³, JI Yimu^{1,2}

(1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 2. Jiangsu HPC and Intelligent Processing Engineer Research Center, Nanjing 210023, China; 3. Jiangsu Yuyue Medical Equipment and Supply Co., Ltd., Zhenjiang 212300, China)

Abstract: Diabetes is a common chronic disease, and it is very important to control blood sugar for preventing diabetes. However, the uncertainty of continuous glucose monitoring (CGM) data extraction significantly increases the difficulty of blood glucose prediction. Therefore, this article proposes a new deep learning based blood glucose concentration prediction model, aiming at improving the model's adaptability to sensor extracted data. In this model, the stacked denoising auto encoder (SDAE) is embedded into the structure of the Transformer encoder to achieve reconstruction, denoising, and feature extraction of input data. Then, a mixed position encoding strategy is adopted to replace the original single absolute position encoding embedding, and a lightweight decoder is introduced into the Transformer model to replace the

基金项目: 江苏省科技支撑计划项目 (BE2019740); 江苏省六大人才高峰高层次人才项目 (RJFW-111); 江苏鱼跃医疗设备股份有限公司科技项目 (2022外 017, 2024外 131); 江苏省研究生实践创新计划项目 (SJCX23_0274)。

收稿日期: 2024-03-01; **修订日期:** 2024-05-01

original structurally complex decoder, aggregate feature information from different levels, and obtain local and global features simultaneously. Finally, by constructing an improved SDAE-improved Transformer network for parallel training of CGM data sequences, temporal patterns and complex correlations in the data can be more comprehensively captured, thus improving predictive performance. Experimental results show that the model has achieved significant performance improvement in blood glucose prediction tasks compared to traditional methods, confirming its effectiveness and robustness in processing CGM data.

Key words: continuous glucose monitoring; neural network; stacked denoising auto encoder (SDAE); Transformer; attention mechanism

引 言

糖尿病作为一种常见的慢性疾病,严重威胁着人们的身体健康。长期糖尿病患者容易引发如酮症酸中毒、高渗性昏迷等严重并发症,还有糖尿病足、心血管病变及视网膜疾病等各种慢性并发症^[1]。因此,对于糖尿病的治疗和预防,做好血糖控制至关重要。然而,由于血糖序列的变化具有非常复杂的非线性关系^[2],且数据采集过程中易受各种因素干扰^[3],想要精确地预测血糖浓度的变化是一项极具挑战的任务。

针对糖尿病预测,国内外学者已经构建了多种基于不同数据以及应用场景的模型。现阶段预测模型主要分为基于生理学和基于数据驱动的方法。生理学模型需要考虑人体胰岛素剂量、体力活动和饮食摄入等因素,描述患者葡萄糖代谢过程,通常较复杂;而数据驱动模型更灵活,并且可以获得更加丰富完整的血糖数据^[5]。早在1993年,Kavanagh等^[6]通过对变量进行皮尔逊相关性检验以及逐步进行多元回归预测,计算出HbA1c百分比以实现糖尿病的控制。2010年,Barakat等^[7]提出了一种可理解的支持向量机算法用于糖尿病的诊断,通过引入规则提取证明了该算法在糖尿病预测的准确性。2015年,王延年等^[8]提出了一种融合差分自回归移动平均(Autoregressive integrated moving average, ARIMA)与径向基神经网络(Radial basis function neural network, RBFNN)的结合血糖预测模型,其中ARIMA负责对未来血糖浓度进行预测,RBFNN对其结果进行修正,有效提高了血糖预测的准确性。2016年,Wang等^[9]提出了一种优化的灰色预测模型,使用少量血糖样本进行预测,结果也表明该模型在改善平滑度的同时可实现较高的预测精度。2019年,Li等^[10]利用卷积神经网络(Convolutional neural network, CNN)与循环神经网络(Recurrent neural network, RNN)进行血糖预测,其中CNN用于捕捉多维时间序列的特征,改进后的RNN则用于分析并学习之前的序列数据。2020年,Gu等^[11]利用神经生理编码器通过端到端的训练与任何其他神经网络模型相结合。2021年,Dudukcu等^[12]结合长短期记忆网络(Long short-term memory, LSTM)、波网(WaveNet)和门控循环单元(Gated recurrent unit, GRU)等架构,通过加权决策级组合预测血糖,得到的最低RMSE值达到21.90 mg/dL。2022年,Naveena等^[5]提出了基于混合元启发式的CNN深度特征提取和改进的基于模糊分类器的糖尿病预测模型,将新的多目标优化算法MF-CSA用于改进深度特征提取、糖尿病检测和血糖水平预测过程。2023年,Langarica等^[13]提出了基于元学习的精确个性化建模技术,将其应用于葡萄糖预测,实验结果证明,只需要对极少量数据量进行很少的训练迭代次数就能实现个性化,并且过拟合的风险极低。

根据上述研究,可见早期糖尿病预测主要依赖机器学习模型,随着卷积神经网络、循环神经网络、长短期记忆网络和门控循环单元等^[14]深度学习模型的发展,更多的关注转向了使用这些模型进行血糖预测。尽管这些模型能在一定程度上实现短期预测,但其自身结构的限制使捕捉长期依赖关系变得困

难,容易出现梯度消失或爆炸等问题。而 Transformer 模型摒弃了传统的递归结构^[15],通过在输入数据中添加位置编码嵌入,利用注意力机制,建立编码-解码框架,从而解决了之前深度学习模型存在的问题。因此,本文选择基于 Transformer 架构的模型进行血糖预测。此外,本文对 Transformer 模型进行了改进,采用混合位置编码策略,同时引入轻量级解码器,替代原始结构复杂的解码器,有效结合局部关注和全局关注,进一步提升模型预测精度。考虑到绝大多数传感器数据涵盖复杂的设备运行状态且采集数据存在各种噪声的污染,本文引入了一种新算法,将堆叠式降噪自编码器(Stacked denoising auto encoder, SDAE)与改进的 Transformer 模型相结合。在进行预测之前,该算法对持续血糖监测(Continuous glucose monitoring, CGM)数据进行高级特征提取,获取更有意义的特征表示,为模型提供更丰富的信息输入。本文主要创新点包括:

(1)针对本文的多维时序数据任务对 Transformer 模型进行改进,采用混合位置编码策略替代原来的单一绝对位置编码嵌入,同时引入轻量级解码器,替代原始结构复杂的解码器,提高模型对时序数据的理解能力。

(2)针对传感器采集数据存在的噪声干扰以及血糖数据的复杂性,本文将 SDAE 与改进 Transformer 模型相结合,实现对数据的隐藏含义的学习,通过 SDAE 的逐层堆叠深层网络,逐渐捕捉数据中的潜在结构,提升模型预测的准确性和鲁棒性。

(3)将本文提出的 SDAE-改进 Transformer 模型应用于血糖预测任务中。模型通过对传感器多维时间序列信号进行重构降噪,将提炼的高级隐含特征输入到 Transformer 回归模型中,进行更全面的时序特征的学习,使模型更高效地处理传感器采集数据,从而更好地适应 CGM 系统中的时序血糖预测任务,提高模型的稳定性和训练效果。

1 相关理论

1.1 CGM 数据的血糖预测研究现状

CGM 是血糖监测领域的一项创新技术突破,它不仅使得血糖监测更加便捷和连续化,还为后续的血糖预测和治疗提供了更为丰富和准确的数据基础。血糖稳态对于糖尿病患者的健康至关重要,而连续血糖监测信号的实时性和连续性为准确预测血糖水平提供了技术支持。

传统的基于生理学方法的血糖预测建模常常受限于数据采集的不足、模型复杂性和个体差异等多种因素。这些限制导致了生理建模在实际应用中的局限,难以满足临床上对于准确预测和个性化治疗的需求。因此,国内外研究者针对 CGM 信号数据提出了多种数据驱动方法,包括自回归(Autoregressive, AR)模型、支持向量机、神经网络等。例如,Rabby 等^[16]提出了一种基于堆叠 LSTM 的深度循环神经网络架构模型,通过卡尔曼平滑技术降低传感器误差,保证 CGM 数据的准确性,以提高血糖预测精度。Hamdi 等^[17]采用支持向量机和差分进化算法预测 1 型糖尿病的血糖值,结果表明基于差分进化算法的支持向量机模型具有较高的预测精度。此外,Jaloli 等^[18]构建了结合 CNN 和 LSTM 的神经网络模型,解决了手动特征工程的限制,并在真实患者数据上证实了其有效性。Yang 等^[19]提出了一种带有模型阶数的 ARIMA 模型,针对 CGM 数据的非平稳性,该辨识算法利用 Akaike 信息准则和最小二乘估计自适应地辨识模型阶数,同时估计相应的参数,实现了较优的性能。

针对上述现有的血糖预测方法,无论是机器学习预测模型还是深度学习预测模型,都只是在原数据基础上改进模型和算法以作出良好的预测效果。然而,CGM 系统提供的数据作为血糖管理的重要工具,在实际应用中仍面临着诸多挑战,其中最显著的是信号的稳定性和准确性问题。受传感器的物理化学特性、天气环境和时间因素等多方面影响,CGM 信号不可避免地受到噪声干扰,以致信号质量

不稳定^[20]。此外,设备故障、探头污染、病人无意识干扰以及情绪变化等因素^[21]对CGM信号的影响也不可忽视。这些问题不仅影响了血糖监测的可靠性,还可能对后续的血糖预测和治疗产生误导性影响,增加了糖尿病管理的难度和风险。此外,对于CGM系统采集的众多血糖数据中,时变性也是一个常见的特征^[22],由于动态监测数据易受人体生理状态、饮食和运动等因素随时的状态而变化,不同时间点的各数据水平可能有所不同,甚至在同一时间点也可能存在波动。因此,对CGM信号数据进行平滑以及降噪处理是非常有必要的。

本文将通过小波分解处理血糖信号数据,并结合SDAE和改进的Transformer模型构建深度学习神经网络模型,以解决CGM信号数据的噪声干扰和时变性问题。SDAE算法将对传感器多维时间序列特征信号进行重构降噪,并将提取出的高级特征输入Transformer回归模型中,实现对时序特征的学习,从而更好地完成CGM数据的时序血糖预测任务。

1.2 Transformer模型原理

长时间历史特征信息的有效记忆对于准确预测CGM系统的血糖水平至关重要,然而传统的循环神经网络和卷积神经网络在处理长序列时可能会面临难以捕捉长距离依赖关系的挑战^[23]。相比之下,Transformer抛弃了递归结构,主要通过自注意力机制,在不同时间步之间灵活地分配注意力权重,从而更好地对序列中的每个元素进行全局建模,并在各个元素之间建立联系^[24]。

Transformer是由Google在2017年提出的一种引入了自注意力机制的深度学习模型,尤其在自然语言处理任务表现出色^[25]。该模型由 N 个相同且前后连接的编码器(Encoder)和解码器(Decoder)两部分组成。其中,编码器负责将输入序列转换为特征表示,而解码器则利用这些特征表示生成输出序列。每个编码器层中分别包括多头自注意力层、残差连接层、标准化层和前馈神经网络层。每个解码器层中在解码器的结构基础上增加了一层多头注意力层。其中,残差连接和层归一化技术主要用于解决梯度爆炸,保证模型的稳定性。这种分层结构使得Transformer模型能够学习不同层次的特征,从而保证模型的泛化能力^[26]。

Bahdanau等^[27]指出Transformer模型与序列到序列模型的区别在于Transformer模型主要利用其核心的自注意力机制进行序列的全局依赖关系捕获,给每个输入序列中的各个位置分配一个权重,最后输出的是加权处理后的位置向量。它的注意力机制具体工作内容包括3个步骤:(1)计算注意力权重,评估每个位置与其他位置之间的关联程度;(2)将每个位置向量与注意力权重相乘,求出加权和向量;(3)进行线性变换,得到最终输出向量。通过多层堆叠架构,构建Transformer模型。

然而,Transformer模型的结构是针对自然语言处理问题提出的,对于处理多维时间序列数据的任务,该模型架构还没有得到很好的研究。但是,使用多头注意力机制处理时序预测的有效建模研究也在逐渐增多,如Hasan等^[28]考虑到Transformer在时序任务的研究不足,提出了一种基于多头注意力的LSTM序列到序列模型,实现对时序信息的学习。因此,本文将在Transformer原始结构的基础上进行改进,利用其自注意力机制实现对多维时间序列数据的预测。

2 模型架构

本研究提出了一种SDAE-改进Transformer血糖浓度预测模型,旨在提高模型对传感器提取数据的适应性。该模型将SDAE算法融入改进后的Transformer网络结构中,利用整合后的网络更全面地捕捉CGM数据中的时序模式和复杂关联,提高模型的预测性能。

模型整体框架主要由数据处理层、SDAE层和Transformer层等组成。数据处理层用于接受模型输入数据,并进行特征选择、小波分解、数据标准化以及数据切片处理;SDAE层对预处理后的数据进行进一步的高级特征提取和降噪;Transformer层对SDAE层输出的特征进行逐层处理,生成预测结果。本文提出的SDAE-改进Transformer血糖预测模型预测流程图如图1所示。

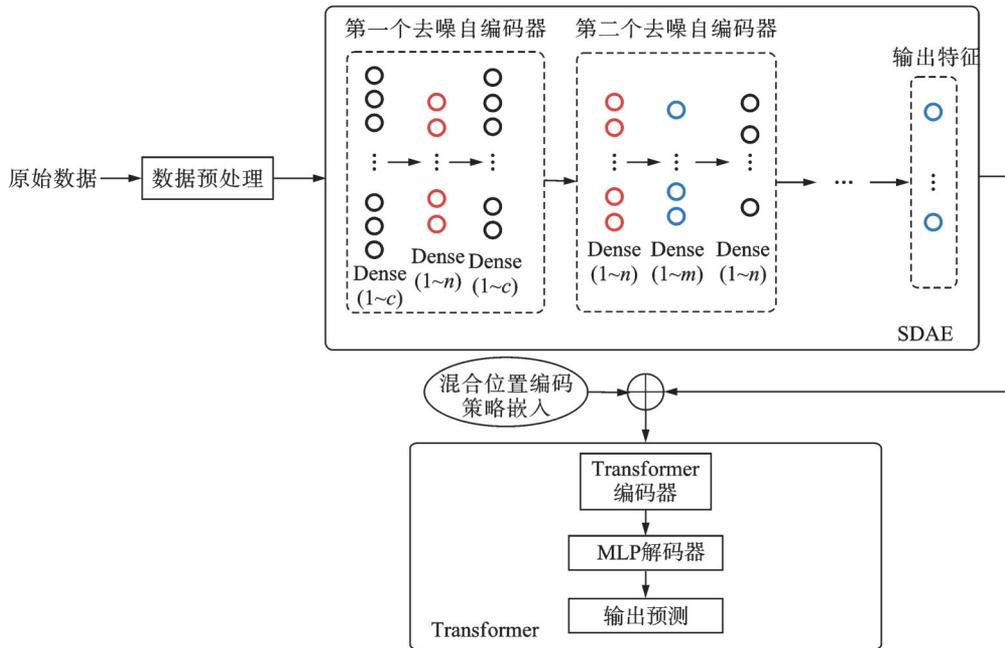


图1 SDAE-改进 Transformer 模型预测流程图

Fig.1 Flow chart of SDAE-improved Transformer model prediction

2.1 数据预处理

本研究所用数据集包括 103 名 2 型糖尿病受试者 7~8 天的传感器数据,删除年龄偏小的 5 例受试以及年龄偏大的 3 例受试,选取剩下的且数据完整的 89 例受试患者数据,总计 973 233 项数据。数据特征包括使用者 id、设备 id、传感器的运行时间、工作电流信号值、空白电流信号值、体表温度、抽血采血血糖值、校准值、统计学处理后的工作电流、未经校准的血糖仪的读数和 CGM 浓度这 11 个特征组成的时间序列数据集 $\{user_id, device_id, itime, iw, ib, t, cno, bg, base_ref, Iw_net, STD\}$ 。

(1) 时间戳特征提取以及特征选择

在特征工程中,首先对时间戳特征进行提取和选择。本文通过单变量特征提取方法,从时间戳字符串中提取年、月、日、时和分等时序特征,并分别以小时和分钟为单位计算了每个受试者佩戴设备的时长。最终选择了原始时间戳中的小时部分“work_hour”和以分钟为单位的佩戴持续时间“minute_length”作为新的时序特征,以便模型更好地理解数据的时间维度动态,提高模型对时间相关特征的敏感度。因此,数据集在经过这一步特征处理后共包含 13 个特征。

此外,为了更深层次了解现有特征间的关系性质,还需要使用特征筛选算法进行特征选择,提高模型学习效率。

首先采用 Pearson 相关系数分析方法,评估每个特征与 CGM 浓度标签 STD 之间的关联性。Pearson 相关性分析是一种常用的统计方法,用于衡量两个变量之间的线性关系强度和方向。它基于协方差的概念,通过计算两个变量的协方差除以它们各自的标准差的乘积,得到一个范围在 -1 到 1 之间的相关系数,提供关于变量之间关系的线索。Pearson 系数的计算方法为

$$s_{A,B} = \frac{\text{cov}(A, B)}{\sigma_A \sigma_B} = \frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^n (A_i - \bar{A})^2} \sqrt{\sum_{i=1}^n (B_i - \bar{B})^2}} \quad (1)$$

式中 s 的值介于 -1 和 1 之间。根据其绝对值,可以判断相关特征相关性的强度:绝对值大于 0.7 为高度相关,介于 0.4 到 0.7 之间为中度相关,在 0.2 与 0.4 之间为弱相关,小于 0.2 为几乎不相关。

其次,采用随机森林模型对每个特征与标签的重要性进行排序。随机森林是一种基于决策树的集成学习算法,旨在将多个决策树组合成一个更强大的模型,其中每个决策树都只使用一部分特征和数据样本来建模,以避免过拟合^[29]。通过构建多个决策树可以评估每个特征对模型的重要性,从而更深入地理解数据的特征。

最后综合考虑皮尔森系数相关性分析和随机森林特征重要性排序的结果,进行特征选择。具体数据系数结果和特征重要性排序如表1所示。根据表1的相关性分析,可见cno和base_ref与血糖浓度之间存在高度相关性。其中,cno为受试者的定期指尖血糖测量值,base_ref为Kellet算法根据最后一次指尖采血生成的未经校准的血糖值,而CGM浓度STD是由Kellet算法根据最后一次指尖采血生成且经过校准的血糖值。由于引入过度相关的特征可能会导致共线性问题并增加了模型解释的难度,因此这两个高度相关特征不考虑作为模型训练的特征输入。而minute_length、ib、device_id、user_id和bg与血糖标签几乎没有明显的线性关系,表明它们对血糖浓度的影响较小。然而缺少线性关系并不代表不存在非线性关系。因此,在结合随机森林模型的特征重要性排序后,最终选择 $\{iw, ib, t, Iw_net, work_hour, minute_length\}$ 这6个特征作为模型训练的输入,以更好地捕捉血糖浓度的变化模式。

(2) 小波包分解

尽管数据采集过程已经经过严格的操作,但由于外在设备的特征和未知因素的干扰,例如信号中断,传感器灵敏度衰减等原因,有时会导致传感器收集的数据出现较大波动,甚至出现偏差和延迟。因此,为确保数据质量和分析结果的准确性,对血糖传感信号数据进行了降噪处理。

本文采用小波包分解信号处理技术,它能够将信号分解成一系列不同尺度的成分,展示信号的局部和整体特征^[30],可在不同频率范围内对数据信号进行详细分析。小波函数作为分解工具,是一组有限长度的波形,在不同的尺度和频率对信号进行分解^[31],因此选择适合的小波函数直接影响到分解的效果。综合各函数特性和本研究数据,选用多贝西小波(Daubechies wavelet 6,db6)函数对血糖传感信号数据进行离散小波变换,将血糖传感信号数据分解到第三层,并获取各频率小波系数,随后,对各频进行软阈值处理,将系数小于预定阈值的部分设为零,而保留大于阈值的系数。最后,利用小波信号的重构原理,对不同细节系数的信号进行逆变换,重构去噪后的信号,得到平滑血糖序列。图2展示了部分

表1 系数结果以及重要性排序

Table 1 Ranking of coefficient results and importance

相关性强弱	特征	Pearson系数	特征重要性排序
高度相关	cno	1.000	1
	base_ref	0.810	2
中度相关	Iw_net	0.650	3
	iw	0.650	4
弱相关	work_hour	0.268	6
	t	0.255	8
几乎不相关	minute_length	0.125	5
	ib	0.107	7
	device_id	0.076	10
	user_id	0.075	11
	bg	0.023	9

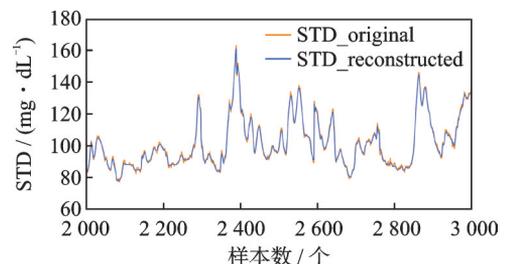


图2 CGM数据经db6小波三层分解重构结果
Fig.2 Reconstruction results of CGM data through db6 wavelet three-layer decomposition

经过 db6 小波三层分解处理和重构的 CGM 血糖传感信号数据段的结果。

(3) 数据标准化

考虑到各项特征度量单位不同,且量纲不一致,为了更好地训练模型,需要对数据进行标准化处理。标准化本质上是通过对线性变换将数据映射到具有特定统计特征的标准分布上,确保不同特征的尺度和分布在训练过程中更一致,并且此种变换不会影响数据序列的原始波动信息。在对训练集和测试集中的各通道数据进行 z -score 标准化方法,以防测试集数据泄露,其公式可表示为

$$y_i = \frac{x_i - \text{mean}}{\text{std}} \quad (2)$$

式中: i 为序列样本点编号,mean为序列中的均值,std为序列中的标准差, y_i 为归一化后的数据。

(4) 数据切片

考虑到在时序数据中不同时刻的数据具有依赖性,为了更有效地学习这种依赖关系,通常采用数据滑窗法对时序数据进行切分,创建序列样本。在训练集切片过程中,假设使用大小为 m 的滑窗对数据集进行划分,随后使用大小为 n 的滑窗对 m 后面的数据进行划分,然后将两个数据窗口同时向后滑动一个数据长度,以此类推。测试集也是相同的切片过程。

经过上述切片处理完的时间序列数据仍不能直接用于神经网络模型的训练。在时序预测任务中,神经网络模型要求时序数据以三维张量数据矩阵的形式输入,包含划分好的数据样本,每个样本时间步长以及输入数据特征维度。因此,还需要将切分好的子序列对转换成监督型学习数据。本文将第 1 个时间点至第 m 个时间点的血糖值作为输入序列 X ,将第 $m+n$ 个时间点的血糖值作为目标输出 Y ,以此类推,转为监督型数据。图 3 为滑动窗口切片数据的示例图。

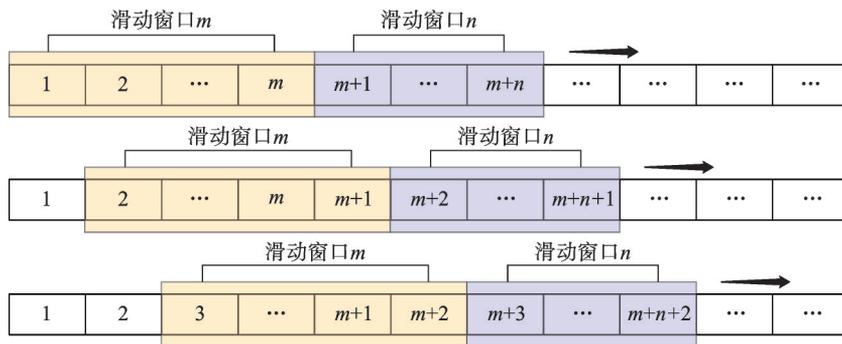


图 3 滑动窗口切片数据示例图

Fig.3 Example image of sliding window slicing data

2.2 堆叠降噪自编码器

CGM 采集数据的各特征间存在复杂的非线性关系,这些特征关系难以通过传统的神经网络模型进行学习。而 SDAE 可以提取数据样本中的隐藏含义,通过逐层堆叠深层网络,智能地学习网络数据的分布规则,并结合各个编码层的系数惩罚和重构误差,对高维数据进行高级特征提取,逐渐捕捉数据中的潜在结构,为后续任务提供更有意义的特征表示^[32]。

SDAE 算法由多个去噪自编码器(Denoising auto encoder, DAE)组成,这些降噪自编码器是对称的神经网络,以非监督的方式学习一些更抽象的特征^[33]。在训练过程中,会在输入数据添加一些噪声,然后通过自编码器的解码器部分去除这些噪声,使得模型在训练过程对有噪声的数据不敏感,保证模型的鲁棒性。

假设输入的数据样本为 x , 添加随机噪声得到含有噪声的信号 \tilde{x} , 在编码部分, 根据从输入层到隐藏层的激活函数 ReLU 函数, 得到如下的隐层向量 h_i 为

$$h_i = f(W^{(en)}\tilde{x} + b^{(en)}) \tag{3}$$

式中: $W^{(en)} \in \mathbf{R}^{L \times M}$ 为编码层的权重, $b^{(en)} \in \mathbf{R}^L$ 为编码层的偏置, $h_i \in \mathbf{R}^L$ 为 L 维的特征, L 为输入的数据样本 x 的维度。解码器中, 模型将从隐藏层的特征向量中重建原始信号 $z_i \in \mathbf{R}^M$, 重建信号 z_i 的计算方式为

$$z_i = f(W^{(de)}h_i + b^{(de)}) \tag{4}$$

最后, 通过如下最小化损失函数, 寻找最优网络参数 $\theta_{SDAE} = \{W^{(en)}, W^{(de)}, b^{(en)}, b^{(de)}\}$, 使重建的信号尽可能接近原始信号, 提取输入的数据样本 x 的深度特征表达为

$$\text{loss}(\theta_{SDAE}) = \frac{1}{2n} \sum_{i=1}^n \|x_i - z_i\|^2 + \gamma_1 \left(\|W^{(en)}\|_F^2 + \|W^{(de)}\|_F^2 \right) \tag{5}$$

式中: γ_1 表示正则化项的系数, $\|\cdot\|^2$ 表示向量的 L2 范数, $\|\cdot\|_F^2$ 表示矩阵的 F 范数。对于其中每层 DAE, 分别将上一个隐藏层的输出作为下一个隐藏层的输入, 然后将多个隐藏层堆叠起来, 之后移除掉堆栈式自编码器的重建层, 获取最后得到的特征。每层 DAE 网络的具体结构如图 4 所示。

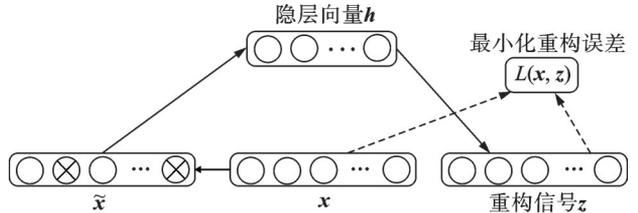


图4 DAE 的网络结构

Fig.4 Network structure of DAE

2.3 改进 Transformer 模型

由于传统的 Transformer 模型是针对自然语言处理问题提出的, 因此本文将在其结构的基础上进行改进, 以适应多维时间序列数据的处理需求。

在图 5 所示的结构中, 本文对 Transformer 模型进行了结构调整, 保留了原始编码器的网络结构, 并引入轻量级的 MLP 层构建解码器, 以更好地适应时序数据的输出。数据输入以后, 首先进行位置编码, 然后通过编码器内部的注意力机制进行特征提取, 最终由轻量级的 MLP 解码器将数据维度转化为一维进行输出。相较于原始 Transformer 模型, 改进后的模型仅需要将数据传入编码器, 无需再进行解码器中的交互注意力计算。这样的结构调整既能利用原始模型编码器中的自注意力机制, 同时又解决了 Transformer 解码器自回归累计误码问题^[34], 降低了模型的复杂度, 并提高了其在时序数据处理任务中的性能和效率。

(1) 混合位置编码策略

位置编码层在模型中起着重要作用, 它通过嵌入位置信息到输入向量中, 使得模型能够捕捉序列数据的时间顺序信息, 并在注意力计

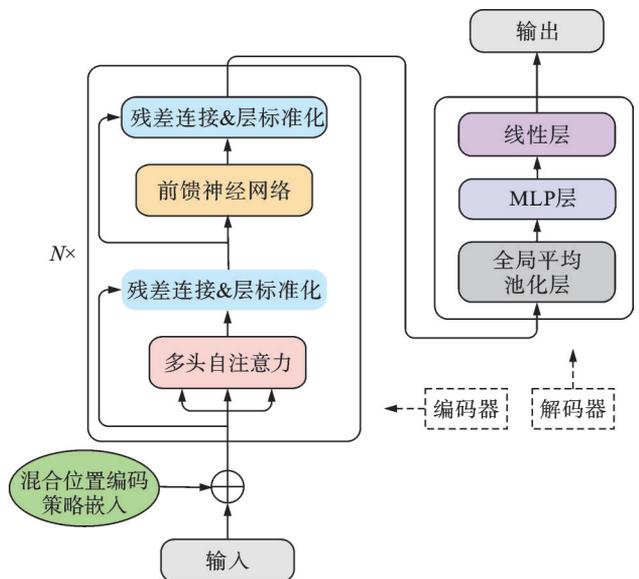


图5 改进的 Transformer 模型架构

Fig.5 Architecture of improved Transformer model

算中避免丢失位置信息。常见的位置编码方式包括绝对位置编码和可学习参数位置编码。绝对位置编码采用正弦或余弦函数在嵌入层添加不同频率表示各位置的位置信息,有助于保持数据原有结构^[35];可学习参数位置编码通过将位置编码参数作为模型的可训练权重,使模型能够在训练过程中学习适应不同任务或数据集中的位置信息,具有更大的灵活性^[36]。

因此,本研究采用混合策略,同时引入了可学习的位置编码和传统的绝对位置编码方式,以允许模型在训练过程中动态地适应不同特征序列的位置信息。假设输入到改进 Transformer 模型的特征向量矩阵为 $X \in \mathbb{R}^{N \times d}$,其中 d 表示特征向量的维度,则绝对位置编码计算方法可以表示为

$$\begin{cases} PE[i, 2j] = \sin(i/10\,000^{2j/d}) \\ PE[i, 2j + 1] = \cos(i/10\,000^{2j/d}) \end{cases} \quad (6)$$

式中:PE 表示位置编码矩阵, i 表示位置编码的位置, j 表示编码向量的维度索引。

(2) 多头注意力机制

注意力机制用于序列数据进行处理时,将输入向量映射到查询矩阵 Q 、键矩阵 K 和值矩阵 V 这 3 个空间向量中,通过查询向量和打分函数计算各个特征的注意力权重^[37]。假设经过位置编码后的矩阵为 $X' \in \mathbb{R}^{N \times d}$, Q 、 K 和 V 矩阵由权重矩阵的乘法实现,计算方式为

$$Q = X'W_Q, K = X'W_K, V = X'W_V \quad (7)$$

式中: W_Q 、 W_K 和 W_V 为线性变换的权重矩阵, Q 、 K 和 V 矩阵每一行维度设为 d_k 。

通过查询向量 Q 和键向量 K 之间的点积进行缩放计算得到特征矩阵的注意力权重,计算公式为

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} \cdot V\right) \quad (8)$$

式中: d_k 为缩放因子;softmax(\cdot) 表示 softmax 映射,通过 softmax 函数进行归一化得到注意力权重。

多头注意力机制是在注意力机制的基础上引入了多个注意力头,以增强模型对不同特征的学习能力,其结构如图 6 所示。每个注意力头有自己的查询矩阵 W_i^Q 、键矩阵 W_i^K 和值矩阵 W_i^V ,计算注意力权重并进行拼接操作,得到最终的多头注意力输出。这使得模型能够并行处理不同方面的信息,提高对关键特征的识别和关注度。多头注意力机制的计算方式为

MultiHead(Q, K, V) =

$$\text{concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^o \quad (9)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (10)$$

式中:Concat 为拼接函数, h 为注意力头的数量, W_i^Q 、 W_i^K 和 W_i^V 分别为 Q 、 K 和 V 第 i 头的线性变换权重矩阵, W^o 为拼接以后线性变换权重矩阵。

(3) 残差归一化层

残差归一化层通过残差连接和层标准化对数据进行处理。残差连接是由 He 等^[38]为了缓解网络退化问题而提出的,通过在多头注意力机制输出向量的基础上加入输入矩阵即位置编码的输出矩阵 X' ,避免模型出现梯度消失的问题,其公式为

$$y = X' + \text{Multihead}(Q, K, V) \quad (11)$$

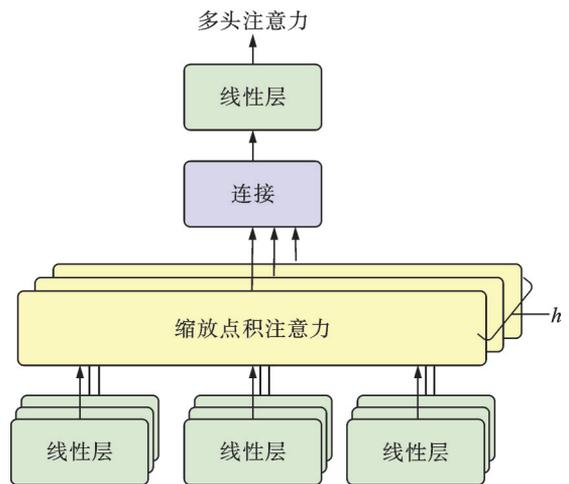


图 6 多头注意力机制

Fig.6 Multi head attention mechanism

层标准化的作用是将每层特征在输入到激活函数前进行归一化,使得每层的特征分布更接近标准正态分布,从而减少数据的偏差,加速模型的收敛,具体计算过程可表示为

$$\text{LayerNorm}(x) = \gamma * \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (12)$$

式中: γ 和 β 为网络参数;“*”表示向量点积; μ 和 σ^2 分别为输入的均值和方差; ϵ 为一个极小值常数,用于避免除以零。

(4) 前馈神经

为了在注意力机制输出信息的基础上加上非线性表达,在Transformer模型中,每个多头注意力的输出通过残差连接后,会进入一个前馈神经网络层。该前馈层由两层全连接层组成,激活函数为ReLU,可表示为

$$\text{FFN}(X) = \text{ReLU}(XW_1 + b_1)W_2 + b_2 \quad (13)$$

式中: X 为多头注意力的输出; W_1 和 W_2 为权重参数矩阵; b_1 和 b_2 为偏置向量。

以上的多头自注意力层、残差连接层、层标准化和前馈网络层等模块组合构成一个编码器子模块。为了强化网络拟合效果,通常会堆叠多个这样的编码器层,构建整个编码器结构,从而提高模型整体的表达能力和学习深层特征的能力。

(5) 轻量级MLP解码器

本文采用轻量级MLP层构建解码器,整合来自不同层的信息,有效结合局部关注和全局关注,同时避免了计算要求高且结构复杂的原解码器。该解码器以编码器提取的多级特征为输入,具体包括以下3个步骤。

① 全局平均池化

该层进行平均池化操作,将编码器的输出序列转换成一个全局向量,捕获整个特征的全局信息,将每个特征维度的信息转化为一个平均的向量,得到一个具有固定长度的向量,增强模型对整体上下文的感知。该向量是整个时序数据的全局信息的汇总,其中每个元素代表相应特征维度上的平均值,公式为

$$y_i = \frac{1}{L} \sum_{j=1}^L X \quad (14)$$

式中: y_i 为每个特征维度的全局平均池化的结果, X 为输入序列, L 为序列数据长度。

② 特征融合

该层任务是对全局平均池化层提取的各维度特征平均值进行通道数的统一操作。主要通过MLP(多层感知器)对每个通道的平均值进行非线性变换和特征选择,以实现特征融合,提高模型对时序数据的抽象能力,增强整体性能。具体公式为

$$Z = \text{Dense}_{C_i, C}(y_i) \quad (15)$$

式中: $\text{Dense}_{C_i, C}$ 为特征融合过程, C_i 为输入通道数, C 为通道数统一后的输出通道数。

③ 预测输出

该层使用全连接层将融合后的特征映射到一个标量,得到预测输出,具体公式为

$$M = \text{Linear}(Z) \quad (16)$$

式中:Linear为线性层, M 为最终预测结果。

2.4 SDAE-改进 Transformer 模型

尽管改进 Transformer 的模型在处理时序数据任务中已经取得显著进展,不再受限于距离,能够灵活地访问历史数据并捕捉数据中的潜在模式和规律。然而,对于复杂的血糖信号预测任务,该模型易受 CGM 数据的时变性和噪声等因素的影响,导致血糖预测精度降低。为此,本研究提出了一种 SDAE-改进 Transformer 血糖浓度预测模型,旨在增强模型对传感器提取数据的适应性。在改进 Transformer 模型中引入了堆叠式降噪自编码器,以更有效地揭示数据样本的潜在结构。

首先,采用 SDAE 算法对原始特征序列进行特征提炼,有针对性地捕捉潜在的数据表征,同时减弱了噪声的影响,随后通过改进 Transformer 模型,对经过 SDAE 处理的特征序列进行更高层次的表征学习。该模型通过整合 SDAE 和 Transformer 两种网络结构,实现更全面的捕捉 CGM 数据中的时序模式和复杂关联,提高预测性能。

基于 SDAE-改进 Transformer 的血糖预测流程如下:

(1) 获取血糖时序数据。利用 CGM 连续采集受试者的传感器数据和血糖序列。

(2) 对数据进行预处理。首先对血糖信号序列进行小波包分解,其次对所有数据进行标准化处理,最后对数据进行切片,并将时序数据转换为三维张量数据矩阵的形式,为后续特征提取和学习做准备。

(3) 将预处理后的数据加入噪声,使用 SDAE 对含有噪声的输入进行多层的编码和解码操作,提炼隐含在原始数据里的高级特征,针对性地捕捉潜在的数据表征,同时减弱噪声的影响。

(4) 使用改进 Transformer 模型对 SDAE 输出的特征序列进行逐层处理,最后通过线性层输出预测结果。

(5) 对 SDAE-改进 Transformer 模型预测结果进行评价指标的计算,分析验证本文所提方法的精确性。

3 实验与分析

3.1 实验数据

2023年7月至9月鱼跃集团在医院共招募103名2型糖尿病受试者,包括46名女性和57名男性。临床数据采集期间,使用安耐糖 Anytime CT3(中国国家药品监督管理局(NMPA)三类医疗器械注册证获批)持续进行7~8天的连续监测。CGM 设备均佩戴在受试者的左或右侧腹部,并由专业看护人员保证设备和数据传输工作的稳定性。受试患者期间遵循饮食、睡眠和活动等方面的临床专业建议,并进行每天不少于6次指尖血采集。研究期间的数据采集主要包括患者和设备静态信息;CT3 传感器每3 min 进行一次采样,共采集1 093 521个传感器采样点;指尖血采血标准数据,共计4 926采血点数据。表2列出了上述各项特征。

3.2 实验配置

本实验使用的计算机配置为 AMD Ryzen 5 处理器(八核)、八核 GPU、16 GB 内存。在 Windows 11 操作系统,使用 Python 3.6 编程语言,PyCharm 作为编程环境。在实验过程中,使用了 Pandas、NumPy、Matplotlib 等库,以

表2 本数据集各特征

Table 2 Characteristics of this dataset	
特征	描述
user_id	使用者 id
device_id	设备 id
itime	传感器的运行时间
iw	工作电流信号值
ib	空白电流信号值
t	体表温度
cno	抽血采血血糖值
bg	校准值
Iw_net	统计学处理后的工作电流
base_ref	未经校准的血糖仪的读数
STD	CGM 浓度

及 Tensorflow2.6.2 深度学习框架,搭建本文提出的血糖预测模型和其他对照组模型框架并通过 GPU 加速计算以提高计算效率。

本节提出的 SDAE-改进 Transformer 血糖预测模型主要参数设置以及模型训练过程中相关超参数的设置如表 3 所示。其中,在 SDAE 层中,用了 ReLU 作为隐藏层的激活函数,并采用 Glorot 均匀初始化来初始化权重,以避免梯度消失或梯度爆炸问题,从而促进模型训练的稳定性。在 MLP 解码器层中,隐藏层和输出层同样采用 ReLU 作为激活函数,以更好地学习特征的非线性特性,权重初始化也使用 Glorot 均匀初始化,保持和 SDAE 层一致,确保模型整体的稳定性和良好的训练效果。

在模型的训练和预测过程中,使用网格搜索及交叉验证的方法为模型选择最优参数。选用了均方误差 (Mean square error, MSE) 作为损失函数,以及 Adam 优化器来动态调整学习率,其初始学习率设置为 5×10^{-3} 。为防止模型训练过程中的过度拟合,引入了 L2 正则化,并对数据进行 Z-score 归一化处理,以提高模型的稳定性和收敛速度。每次训练批次包含 32 个样

本,训练轮数设定为 100 次。此外,为进一步提升模型性能和稳定性,引入了 ModelCheckpoint 回调函数和 EarlyStopping 回调函数。ModelCheckpoint 回调函数用于监测验证集上的 MAE,并在发现更低的 MAE 时保存具有最佳性能模型;而 EarlyStopping 回调函数实现在连续 10 次迭代中损失函数未改善时提前停止训练。最后,所有实验结果经过 5 次取均值处理,以确保结果的稳定性和可靠性。

3.3 实验评价指标

为了全面评价本文提出的血糖预测模型的性能,采用 3 种主要的预测模型评价指标,分别是均方根误差 (Root mean square error, RMSE)、平均绝对误差 (Mean absolute error, MAE) 和克拉克误差网格分析法 (Clark error grid analysis, EGA),本实验中 RMSE、MAE 的单位均为 mg/dL。

均方根误差计算公式为

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

平均绝对误差计算公式为

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

式中: n 为样本数量, \hat{y}_i 为第 i 个样本实际值, y_i 为第 i 个样本预测值。

克拉克误差网格分析是在 1987 年开发的分析方法,用于衡量患者的实际血糖值和预测血糖值的临床精确度^[39],是一种常用于血糖预测任务评价的方法,它通过将预测值与实际值在二维空间中进行比较,将预测误差分为 A、B、C、D、E 这 5 个区域,用以评估血糖预测方法的准确度,从而更直观地评估预测模型的准确性和临床意义,如图 7 所示。在图 7 中,区域 A 表示预测值和实际值在临床上没有差异,

表 3 模型各区域参数设置

Table 3 Parameter settings for each region of the model

所属区域	参数名	参数设置
SDAE 层	稀疏因子	0.05
	SDAE 网络深度	2(32,16)
Transformer 编码器层	注意力头数	256
	注意力维度	4
	前馈网络深度	4
	编码器网络深度	4
MLP 解码器层	Dropout 率	0.25
	每层神经元数量	128
	Dropout 率	0.25
超参数	学习率	0.000 1
	优化函数	Adam
	损失函数	MSE

预测误差对患者的管理和治疗没有明显影响;区域B表示预测值和实际值之间存在轻微的差异,但对患者的管理和治疗影响较小,可以接受;区域C表示预测值和实际值之间存在较大的差异,可能会对患者的管理和治疗产生一定影响,需要谨慎处理;区域D表示预测值和实际值之间存在严重的差异,可能导致患者的管理和治疗出现错误或不良影响,需要立即采取措施进行修正;区域E表示预测值和实际值之间存在危险的差异,可能会导致患者的健康状况急剧恶化,需要立即采取紧急措施进行处理。通过对预测误差进行分区,使其可以与实际临床结果联系起来,并且可以更直观地评估模型的可靠性和适用性。区域A和B表示模型的预测较为准确,对患者的管理和治疗影响较小,而区域C、D和E则表示需要引起关注,可能需要进一步优化模型或采取其他措施来提高预测的准确性和临床可靠性。

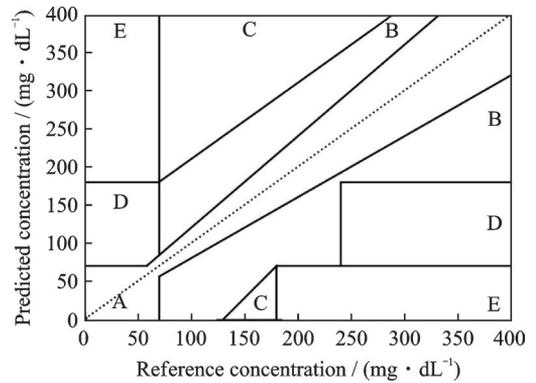


图7 克拉克网络分析图

Fig.7 Clark network analysis diagram

3.4 实验结果分析

为了验证本文所提算法的预测能力及泛化能力,通过调节滑窗大小达到不同输入时长的预测,测试其提前步长分别为5、10、15和20步的预测效果。在实验中,根据血糖控制效果对数据集进行分层抽样,70%受试者的连续序列作为训练集,30%受试者的连续序列作为测试集,其中训练集包括62例受试者,测试集包括27例受试者。表4展示了10名不同受试者提前时长分别是15、30、45和60 min时的RMSE和MAE预测结果分析。图8展示了该受试者经过本文模型分别提前15、30、45和60 min时的部分区域血糖趋势预测曲线。

表4 10名用户不同提前时长血糖预测评价指标结果对比

Table 4 Comparison of blood glucose prediction evaluation indicators for ten users at different times in advance mg/dL

用户	评价 指标	提前预测时长				用户	评价 指标	提前预测时长			
		15 min	30 min	45 min	60 min			15 min	30 min	45 min	60 min
1	RMSE	18.681	19.847	20.007	20.168	6	RMSE	19.958	21.925	22.335	26.651
	MAE	14.214	14.991	15.143	15.219		MAE	14.916	16.571	16.786	20.526
2	RMSE	17.296	18.902	20.452	21.445	7	RMSE	18.823	20.018	20.926	21.980
	MAE	13.023	13.970	15.537	15.062		MAE	14.883	15.968	16.147	17.535
3	RMSE	15.034	16.245	18.694	21.038	8	RMSE	13.781	14.287	15.187	16.568
	MAE	11.271	12.421	14.577	16.605		MAE	10.257	11.246	12.285	13.258
4	RMSE	23.937	25.529	26.095	28.255	9	RMSE	14.826	15.381	16.431	18.6763
	MAE	17.2381	18.330	19.488	20.322		MAE	11.371	12.009	12.730	15.316
5	RMSE	16.301	16.207	17.281	20.002	10	RMSE	15.195	20.927	25.314	23.939
	MAE	13.067	12.501	13.464	14.109		MAE	12.367	16.553	20.428	19.003

为了验证本模型架构各个部分的有效性,以及相较于其他模型的预测性能,对本文模型与 CNN、LSTM、GRU、Transformer 以及改进 Transformer 等方法进行了对比,计算各模型的 RMSE 和 MAE,并绘制了预测误差对比图。8号用户在输入模型序列时长为 15 min 时,上述各模型预测结果对比各项评价指标如表 5 所示。从表 5 可以看出,CNN 模型的 RMSE 和 MAE 值最大,且预测效果最差。这主要由于 CNN 模型容易出现信息丢失和误差累积的问题,在处理时序数据时可能会丢失长期依赖性。相比之下,LSTM 和 GRU 模型在预测效果得到了显著的提升,且两者预测结果相差不大,但它们无法解决梯度消失或梯度爆炸

等问题。传统的 Transformer 模型相较于 GRU 模型,血糖趋势预测结果的 RMSE 降低了 1.275 mg/dL, MAE 降低了 1.091 mg/dL,可见 Transformer 模型的并行化处理优势能够更准确地预测未来血糖水平的变化趋势。在按照本文的研究对 Transformer 进行改进时,模型预测效果得到了进一步提升,其中 RMSE 降低了 2.922 mg/dL,MAE 降低了 2.496 mg/dL。这种改进的 Transformer 模型能更全面地理解时序数据中的信息,其注意力机制和层数的调整使得模型更能够捕捉长期依赖性和复杂关联,从而减少了预测偏差。此外,加入模型中的 SDAE 模块后,进一步降低了 RMSE 和 MAE,提高了模型对数据中潜在模式和规律的理解能力。综上可见,本文提出的 SDAE-改进 Transformer 血糖趋势预测模型的预测效果要优于其他模型,并且每个模型各部分的改进都有助于血糖趋势预测结果的精确性。

由表 6 和图 9 可知,所有模型的预测点均在 A、B 区域内,本模型预测时,A 区域占比 95.74%,B 区域占比 4.26%,达到其中最好的血糖预测效果。可见,本模型预测的结果优于其他模型的预测效果,说明本模型对 CGM 数据序列的并行化训练能更全面地捕捉数据中的时序模式和复杂关联,提高预测性能。

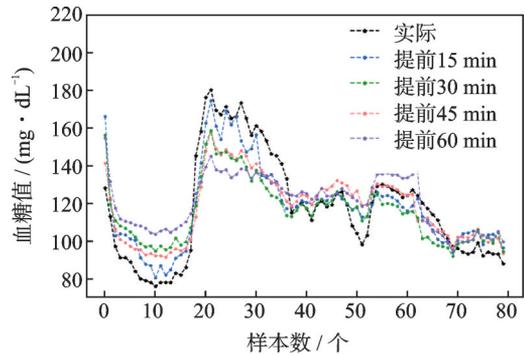


图 8 提前不同时间的部分区域血糖趋势预测曲线

Fig.8 Prediction curves of partial regional blood glucose trends at different times in advance

表 5 各模型预测结果对比各项评价指标

Table 5 Comparison of various evaluation indicators for predicted results of each model

	mg/dL	
模型	RMSE	MAE
CNN	25.746	20.057
LSTM	20.206	15.781
GRU	20.124	15.845
Transformer	18.849	14.754
改进 Transformer	15.927	12.258
SDAE-改进 Transformer	13.781	10.257

表 6 不同模型的 EGA 中 A、B 区域占比

Table 6 Proportion of regions A and B in EGA of different models

模型	区域		%
	A	B	
CNN	77.25	22.75	
LSTM	88.05	11.95	
GRU	87.12	12.88	
Transformer	90.94	9.06	
改进 Transformer	94.93	5.07	
SDAE-改进 Transformer	95.74	4.26	

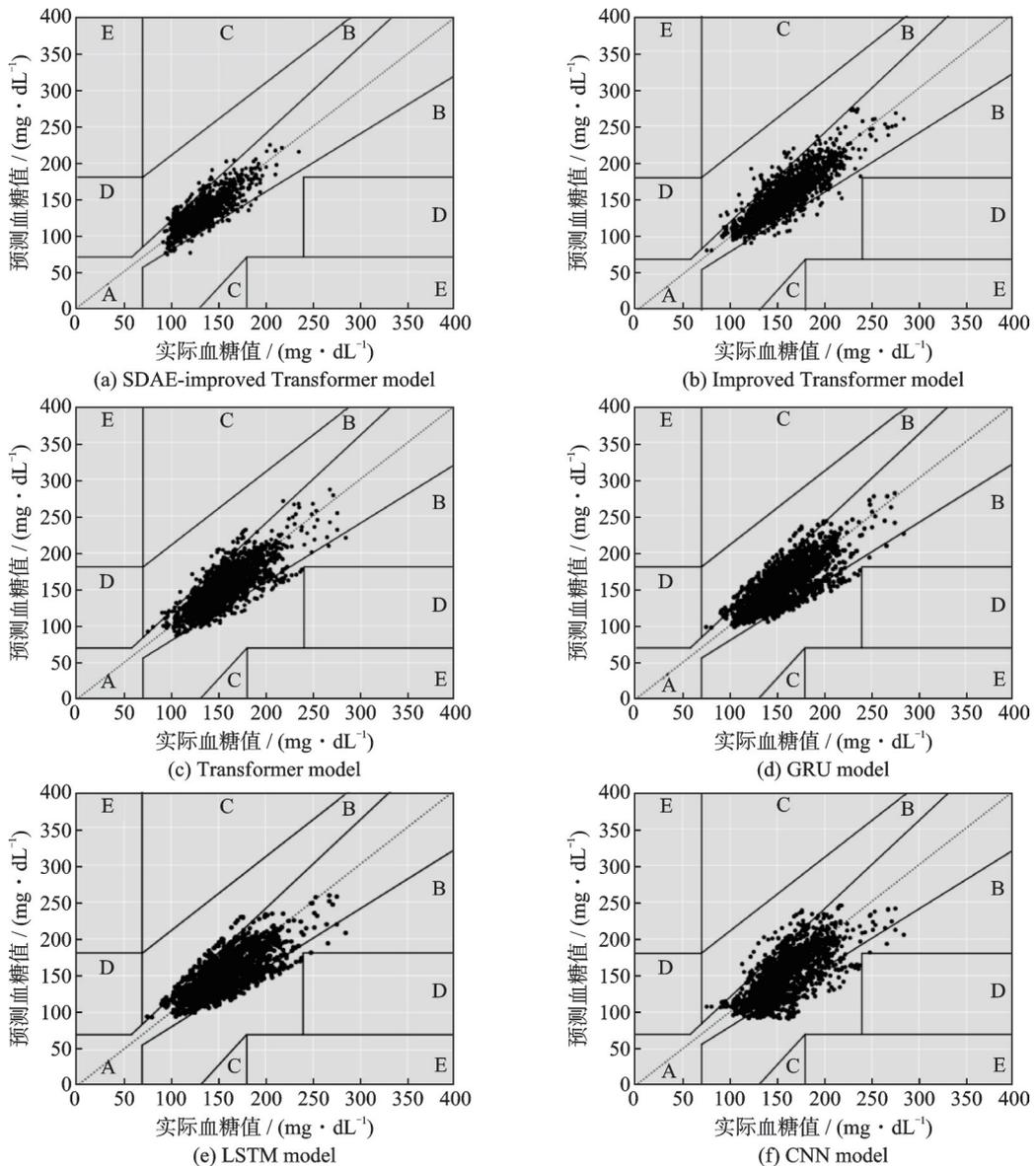


图9 不同模型的 EGA 图

Fig.9 EGA diagrams of different models

4 结束语

本文结合了 CGM 数据的时序特点,提出了一种将 SDAE 与改进 Transformer 相结合的模型。首先,利用 SDAE 算法对 CGM 数据进行高级特征提取,为模型提供更有意义的特征表示;其次,对 Transformer 模型进行了改进,采用混合位置编码策略,并引入轻量级解码器来取代原始结构复杂的解码器,实现局部关注和全局关注的有效结合,进一步提升模型预测精度;最后通过对比试验,验证了本模型在均方根误差和平均绝对误差上的最佳性能,且在克拉克误差网格分析图中落在 A 区域的预测点占比最大,且都落在 A 和 B 区域,进一步验证了本模型架构各个部分的有效性以及相较于其他模型的预测性能。未来,将尝试进一步扩展数据集规模并提升数据质量,以更全面地评估模型的泛化能力,同时也将致力于将该模型应用于实际临床实践中,验证其在实际场景下的有效性和实用性。

参考文献:

- [1] ALBERTI K G M M, ZIMMET P Z. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus provisional report of a WHO consultation[J]. *Diabetic Medicine*, 1998, 15(7): 539-553.
- [2] WU F, WU Q, TAN Y. Workflow scheduling in cloud: A survey[J]. *The Journal of Supercomputing*, 2015, 71(9): 3373-3418.
- [3] GUTIERREZ G J O, SIM K M. A family of heuristics for agent-based elastic cloud bag-of-tasks concurrent scheduling[J]. *Future Generation Computer Systems*, 2013, 29 (7): 1682-1699.
- [4] XIE J, WANG Q. Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-series models[J]. *IEEE Transactions on Biomedical Engineering*, 2020, 67(11): 3101-3124.
- [5] NAVEENA S, BHARATHI A. A new design of diabetes detection and glucose level prediction using moth flame-based crow search deep learning[J]. *Biomedical Signal Processing and Control*, 2022, 77: 103748.
- [6] KAVANAGH D J, GOOLEY S, WILSON P H. Prediction of adherence and control in diabetes[J]. *Journal of Behavioral Medicine*, 1993, 16(5): 509.
- [7] BARAKAT N, BRADLEY A P, BARAKAT M N H. Intelligible support vector machines for diagnosis of diabetes mellitus [J]. *IEEE Transactions on Information Technology in Biomedicine*, 2010, 14(4): 1114-1120.
- [8] 王延年, 雍永强, 贾晓灿, 等. 融合 ARIMA 和 RBFNN 的血糖预测[J]. *计算机工程与设计*, 2015, 36(6): 1652-1656.
WANG Yannian, YONG Yongqiang, JIA Xiaocan, et al. Prediction of blood glucose fusing ARIMA and RBFNN[J]. *Computer Engineering and Design*, 2015, 36(6): 1652-1656.
- [9] WANG Y N, WEI F F, SUN C Q, et al. The research of improved grey GM (1, 1) model to predict the postprandial glucose in type 2 diabetes[J]. *BioMed Research International*, 2016(9): 1-6.
- [10] LI K, DANIELS J, LIU C, et al. Convolutional recurrent neural networks for glucose prediction[J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(2): 603-613.
- [11] GU K, DANG R, PRIOLEAU T. Neural physiological model: A simple module for blood glucose prediction[C]//*Proceedings of 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. [S.l.]: IEEE, 2020: 5476-5481.
- [12] DUDUKCU H V, TASKIRAN M, YILDIRIM T. Blood glucose prediction with deep neural networks using weighted decision level fusion[J]. *Biocybernetics and Biomedical Engineering*, 2021, 41(3): 1208-1223.
- [13] LANGARICA S, RODRIGUEZ-FERNANDEZ M, NUNEZ F, et al. A meta-learning approach to personalized blood glucose prediction in type 1 diabetes[J]. *Control Engineering Practice*, 2023, 135: 105498.
- [14] ZHU T Y, LI K Z, HERRERO P. Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning[J]. *IEEE Transactions on Biomedical Engineering*, 2023, 70(1): 193-204.
- [15] CICHOSZ S L, JENSEN M H, HEJLESEN O. Short-term prediction of future continuous glucose monitoring readings in type 1 diabetes: Development and validation of a neural network regression model[J]. *International Journal of Medical Informatics*, 2021, 151: 104472.
- [16] RABBY M F, TU Y, HOSSEN M I, et al. Stacked LSTM based deep recurrent neural network with Kalman smoothing for blood glucose prediction[J]. *BMC Medical Information Decision Making*, 2021, 21(1): 101.
- [17] HAMDI T, ALI J B, DI COSTANZO V, et al. Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm[J]. *Biocybernetics and Biomedical Engineering*, 2018, 38(2): 362-372.
- [18] JALOLI M, CESCÓN M. Long-term prediction of blood glucose levels in type 1 diabetes using a CNN-LSTM-based deep neural network[J]. *Journal of Diabetes Science Technology*, 2023, 17(6): 1590-1601.
- [19] YANG J, LI L, SHI Y, et al. An ARIMA model with adaptive orders for predicting blood glucose concentrations and hypoglycemia[J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(3): 1251-1260.
- [20] KUSUNOKI Y, KATSUNO T, NAKAE R, et al. Comparison of numerical accuracy of personal and professional continuous glucose monitors[J]. *Journal of the Japan Diabetes Society*, 2015, 58(9): 715-720.
- [21] GINSBERG B H. Factors affecting blood glucose monitoring: Sources of errors in measurement[J]. *Journal of Diabetes Science and Technology*, 2009, 3(4): 903-913.
- [22] ACCIAROLI G, VETTORETTI M, FACCHINETTI A, et al. From two to one per day calibration of dexcom G4 platinum by a time-varying day-specific Bayesian prior[J]. *Diabetes Technology & Therapeutics*, 2016, 18(8): 472-479.

- [23] VALERI J A, COLLINS K M, RAMESH P, et al. Sequence-to-function deep learning frameworks for engineered riboregulators[J]. Nature Communications, 2020, 11(1): 5058-5073.
- [24] CHANDRA A, TÜNNERMANN L, LÖFSTEDT T, et al. Transformer-based deep learning for predicting protein properties in the life sciences[J]. Elife, 2023, 12: 82819.
- [25] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2017-06-12). <https://doi.org/10.48550/arXiv.1706.03762>.
- [26] AGRAWAL P, BANSAL H O, GAUTAM A R, et al. Transformer-based time series prediction of the maximum power point for solar photovoltaic cells[J]. Energy Science & Engineering, 2022, 10(9): 3397-3410.
- [27] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]//Proceedings of International Conference on learning Representations. [S.l.]: [s.n.], 2014.
- [28] HASAN F, HUANG H. MALS-Net: A multi-head attention-based LSTM sequence-to-sequence network for socio-temporal interaction modelling and trajectory prediction[J]. Sensors, 2023, 23(1): 530.
- [29] LIN W W, WU Z M, LIN L X, et al. An ensemble random forest algorithm for insurance big data analysis[J]. IEEE Access, 2017, 5: 16568-16575.
- [30] WANG G, YANG L, LIU M, et al. ECG signal denoising based on deep factor analysis[J]. Biomedical Signal Processing and Control, 2020, 57: 101824.
- [31] ZHANG C, SUN T. Discussion of the influence of multiscale PCA denoising methods with three different features[J]. Sensors, 2022, 22(4): 1604.
- [32] WANG E K, ZHANG X, PAN L. Automatic classification of CAD ECG signals with SDAE and bidirectional long short-term network[J]. IEEE Access, 2019, 7(99): 182873-182880.
- [33] DONG Z H, HOU K, LIU Z Y. A deep-learning-based optimal energy flow method for reliability assessment of integrated energy systems[J]. IEEE Access, 2022, 10: 91092-91102.
- [34] XIE E, WANG W, YU Z, et al. SegFormer: Simple and efficient design for semantic segmentation with Transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
- [35] JIANG K, PENG P, LIAN Y, et al. The encoding method of position embeddings in vision transformer[J]. Journal of Visual Communication and Image Representation, 2022, 89: 103664.
- [36] LI Y, SI S, LI G, et al. Learnable Fourier features for multi-dimensional spatial positional encoding[J]. Advances in Neural Information Processing Systems, 2021, 34: 15816-15829.
- [37] 徐鹤, 郑群力, 谢作玲, 等. 基于知识表示向量的可解释深度学习模型及其疾病预测应用[J]. 数据采集与处理, 2023, 38(4): 777-792.
- XU He, ZHENG Qunli, XIE Zuoling, et al. Interpretable deep learning model based on knowledge representation vectors and its application in disease prediction[J]. Journal of Data Acquisition and Processing, 2023, 38(4): 777-792.
- [38] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2016: 770-778.
- [39] CLARKE W L, COX D, GONDER-FREDERICK L A, et al. Evaluating clinical accuracy of systems for self-monitoring of blood glucose[J]. Diabetes Care, 1987, 10(5): 622-628.

作者简介:



徐鹤(1985-),通信作者,男,博士,教授,研究方向:人工智能与大数据、物联网技术, E-mail: xuhe@njupt.edu.cn。



杨丹丹(2000-),女,硕士研究生,研究方向:人工智能和大数据、深度学习。



刘思行(1991-),男,硕士,工程师,研究方向:人工智能、生物电信号、生物化学信号分析与处理。



季一木(1978-),男,博士,教授,博士生导师,研究方向:高性能计算、大数据和人工智能, E-mail: jiyim@njupt.edu.cn。