

DOI:10.19651/j.cnki.emt.2416033

基于改进 Transformer-BiLSTM 的人体活动识别模型

孙巍伟¹ 毛亦鹏¹ 郑家春² 梁毅玮¹

(1. 北京信息科技大学机电工程学院 北京 100192; 2. 北京遥感设备研究所 北京 100854)

摘要: 针对可穿戴传感器采集的时间序列往往具有维度高、噪声大等缺点导致活动识别方法准确率下降的问题,提出了基于改进 Transformer-BiLSTM 的人体活动识别模型。模型采用了 Transformer 编码器在处理长距离依赖和并行化计算方面的优势来提高序列特征提取的效率;随后将特征传递给添加了跳跃残差连接的双向长短期记忆网络,两次残差连接代替大量卷积层的同时保留了有效信息;提出了一种集成有时间信息编码的注意力层增强了模型的表达能力和对时序数据的理解能力。实验结果表明,该模型在公开数据集上的准确率达到 98.38%,有效提高了人体活动识别的准确率。

关键词: 步态识别;深度学习;Transformer;双向长短期记忆网络;特征融合

中图分类号: TN876 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Human activity recognition model based on improved Transformer-BiLSTM

Sun Weiwei¹ Mao Yipeng¹ Zheng Jiachun² Liang Yiwei¹

(1. School of Mechanical Electrical Engineering, Beijing Information Science & Technology University, Beijing 100192, China;

2. Beijing Remote Sensing Equipment Research Institute, Beijing 100854, China)

Abstract: A human activity recognition model based on an improved Transformer-BiLSTM network is proposed to address the problem of decreased accuracy in activity recognition methods due to the high dimensionality and large noise of time series collected by wearable sensors. The model leverages the advantages of Transformer encoder in handling long-range dependencies and parallelized computations to enhance the efficiency of sequence feature extraction. Subsequently, the features are passed to a bidirectional long short-term memory network with skip residual connections, where two residual connections replace numerous convolutional layers while retaining essential information. Additionally, an attention layer integrated with time information encoding is proposed to enhance the model's expressive power and understanding of temporal data. Experimental results show that the model achieves an accuracy of 98.38% on public datasets, effectively improving the accuracy of human activity recognition.

Keywords: gait recognition; deep learning; Transformer; two-way long short-term memory network; feature fusion

0 引言

人体活动识别(human activity recognition, HAR)通过分析人体活动信号来识别运动行为模式,在智慧医疗领域有着重要的经济价值和科学研究意义^[1]。从广义上讲,基于人工智能的 HAR 系统可以分为两类:基于视觉^[2]和基于传感器^[3]。基于视觉的方法利用图像处理技术分析数据,但受到照明条件、噪声和其他物理干扰的影响。相比之下,基于传感器的方法价格低廉,且不受外部环境的影响。这种差异使得基于传感器的 HAR 系统更大众化,并适用于不同的实际场景,如病人的健康监测、老年人的跌倒检测、运动员的活动跟踪和心电图监测等^[4]。

通常,基于传感器的 HAR 被定义为多变量时间序列分类任务,在空间维度上提取时间帧的特征,并按时间顺序堆叠空间特征。传感器如加速度计(测量加速度)、陀螺仪(检测角速度和方向)等被用于从输入数据中提取有用信息。在提取高维度时间序列数据特征方面,传统的 HAR 方法以机器学习方法为主,需人工提取特征,因此特征的选择对识别的精度具有较大影响。吴海涛等^[5]设计了自动提取传感器数据特征和记忆时序性活动数据的 CNN-GRU 混合神经网络模型,在公开数据集上能达到预期的高准确率;王晓玲^[6]在 7 个数据集上采用欧氏距离为标准进行特征筛选,使用各数据集的特征权重最大的十个特征,作为算法的训练和评估数据。深度学习的一个显著优势是能够将

特征提取过程融入到模型中,使网络能够自动学习到有效的特征。这一优点使得深度学习逐渐替代了高度依赖手动特征选择的传统机器学习方法。虽然学者们在特征工程方面的研究仍在继续,但深度学习的崛起显著简化了这一过程,提升了模型的性能和适用性。在对时间序列的处理上,深度学习可以使用编码器进行原始数据的降维,同时保证原有的时序特征,因而延长了模型可以处理的时间序列长度,Pramannik 等^[7]提出的反向注意力机制能够在学习过程中自我校准,并动态调整学习率;Milon 等^[8]利用多头卷积神经网络和卷积块注意力模块处理视觉数据,利用卷积长短期记忆处理时间敏感的多源传感器信息;魏淇等^[9]采用基于局部特征的时间注意力图卷积步态识别方法,并通过公开数据集进行了验证;陈万志等^[10]提出了一种改进的 GaitSet 步态识别方法并在公开数据集上进行了验证。Sullivan 等^[11]采用了深度特征分析方法,对模型中的特征进行了深入分析并得到对于每个特征的重要性,提高了模型的可解释性,但其准确率只有 83%;张红颖等^[12]提出一种结合残差网络和多级分块结构的步态识别方法,在公开数据集上进行实验,在携带背包或不同衣着条件下的识别率分别达到了 87.5%、82.6%。

综上所述,以上方法对于长序列、高维度的时间序列分类准确率低的问题的解决方法主要以各种卷积神经网络(convolutional neural networks, CNN)的变体为主。但是 CNN 通过堆叠卷积层来实现对时间序列的提取,会导致产生计算量庞大的模型、梯度消失、梯度爆炸等问题,网络将难

以训练和收敛^[13],在实践中存在模型训练难度大等问题。

Transformer 神经网络模型相对于卷积神经网络 CNN 或者长短时记忆网络(long short-term memory, LSTM),由于其独特的自相关结构,能够对不同特征维度之间的关系进行充分地挖掘,使得 Transformer 网络模型相对于其他深度网络,其对特征维度之间的数据关系更为敏感。为解决高维度的长时间序列分类效率低、准确率低等问题,本文提出了一种改进 Transformer-BiLSTM 深度学习模型用于人体活动识别。模型采用了 Transformer 编码器作为特征提取模块,既融合了特征,又通过多头注意力机制高效处理长距离依赖关系;然后将融合后的特征传入改进的双向长短时记忆网络(bidirectional long short-term memory, BiLSTM)网络模块中,能够有效地捕捉序列数据中的时序信息;设计新的网络结构,以跳跃的两次残差连接代替传统卷积层的大量堆叠,有效解决上下采样过程造成的特征丢失,最后引入提出基于时间信息的位置编码注意力机制可以关注时间特征和提供更多的灵活性。

1 改进 Transformer-BiLSTM 模型

1.1 模型结构

研究旨在从原始传感器数据中识别人体活动模式。模型整体结构由数据预处理模块、特征提取模块(Transformer 编码器)、改进的 BiLSTM 模块(包括时间分布层、BiLSTM 层、残差连接)和改进的注意力模块组成,模型结构如图 1 所示。

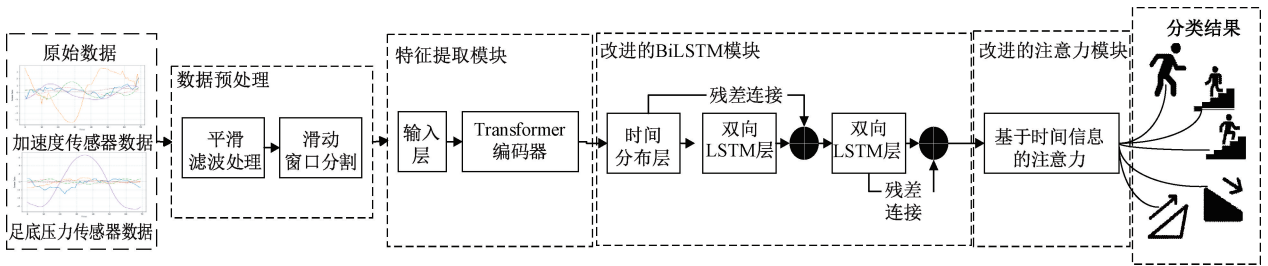


图 1 模型整体结构图

Fig. 1 Overall structure diagram of the model

首先,输入层接收形状为(窗口大小,特征维度)的张量,代表每个样本窗口的时间步数和特征维度。接着是 Transformer 编码器,它包括嵌入层、多头注意力机制和前馈网络,用于捕捉全局信息和局部依赖关系。然后是时间分布层,其作用是将每个时间步的特征映射到 128 维的空间中。接下来是双向长短时记忆网络,其中包含 64 个 LSTM 单元,用于进一步处理数据。残差连接将 Transformer 编码器的输出与第一个 BiLSTM 层的输出相加,有助于信息传播。再次添加残差连接合并两个双向 LSTM 层的输出。在模型的最后阶段,利用改进的位置编码与注意力机制相结合,以更好地处理足底压力数据和身体多部位三轴加速度数据。通过训练过程,模型能够学习

到每个传感器数据的重要性,并根据其重要性分配相应的权重,从而更有效地捕获序列数据中的模式和关系。最后,输出层是一个全连接层,将特征映射到标签的类别上,使用 softmax 激活函数对特征进行分类,得到最终的分类结果。

1.2 基于 Transformer 编码器的特征提取模块

Transformer 采用了解码器-编码器结构,编码器用于将输入序列编码为一个上下文表示,而解码器则使用此上下文表示来生成输出序列,但 Transformer 通过使用多头注意力机制代替了循环神经网络结构,使得整个模型可以并行计算。例如处理文本数据,这些数据本身是有顺序的。将一个文本序列作为输入,并产生另一个文本序列作

为输出。而只有编码器的 Encoder-only 架构则更擅长对内容进行分类,本文所提出模型的任务是基于时间序列数据进行分类,所以本文采用的为该结构,将编码结果直接连接双向 LSTM 层。多头注意力机制通过并行运行多个自注意力层并综合其结果,能够同时捕捉输入序列在不同子空间中的信息,从而增强模型的表达能力,自注意力的计算方法如式(1)。

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

式中: $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别代表 Query、Key、Value 矩阵; d_k 为 $\mathbf{K}$ 矩阵的维度。

多头注意力的公式表达如下:

$$\text{MutiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^0 \quad (2)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (3)$$

式中: i 是注意力头的个数, $(\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V)$ 均为线性变换时的参数矩阵。由于 Concat 后的矩阵实际上并不是有机地融合多个自注意力头的输出,而只是简单地做了矩阵的前后链接,所以在 Concat 后需要通过 $\mathbf{W}^0$ 矩阵再做一次线性变换, $\mathbf{W}^0$ 也是在模型训练阶段一同训练出来的参数矩阵。Transformer 编码器的结构如图 2 所示。

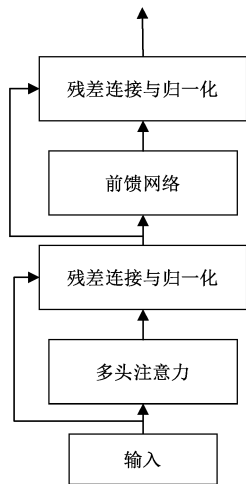


图 2 Transformer 编码器结构图

Fig. 2 Transformer encoder structure diagram

1.3 改进的双向长短期记忆网络

长短期记忆网络(LSTM)之所以能够解决循环神经网络(RNN)的长期依赖问题,是因为 LSTM 引入了门机制用于控制特征的流通和损失。LSTM 引入了 3 个门(输入门 i_t 、遗忘门 f_t 、输出门 O_t)和细胞状态 C_t , LSTM 结构如图 3 所示, X_t 是输入信息, h_t 是基于细胞状态得到的隐藏状态。

从输入开始,LSTM 的第一步要决定从细胞状态中舍弃哪些信息。这一决定由遗忘门 f_t 做出:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

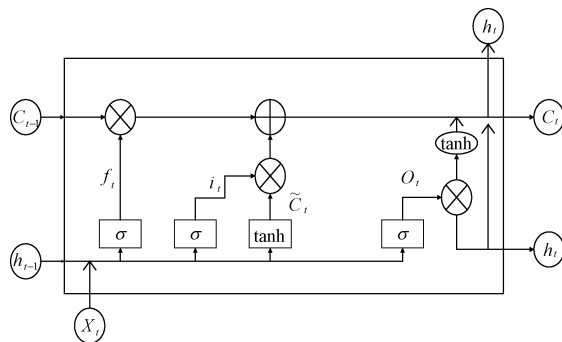


图 3 LSTM 结构图

Fig. 3 LSTM structure diagram

下一步中由输入门确定哪些新信息将被保存在细胞状态中,这一步包含两个部分,分别是 sigmoid 激活函数组成的“输入门”产生的介于 0~1 之间的控制信号 i_t ,用来控制 $\tilde{C}_t$ 输入的程度,和通过一个 tanh 激活函数产生当前时刻的候选细胞状态 $\tilde{C}_t$,这个值将由 i_t 决定添加到细胞状态中的程度。公式如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

第 3 步则是更新细胞状态 C_t , 公式如下:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (7)$$

最后一步是输出门,与输入门相同也包括两部分,第 1 部分,由 sigmoid 组成的“输出门”产生的介于 0~1 之间的控制信号 o_t ;第 2 部分,将最终产生的输出信息 $\tanh(C_t)$ 与控制信号 o_t 相乘,得到最终的输出值 h_t 。输出门控制记忆单元 C_t 对当前输出值 h_t 的影响,即记忆单元中的哪一部分会在时间步 $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ 输出,公式如下:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \times \tanh(C_t) \quad (9)$$

BiLSTM 通过使用两个独立的 LSTM 网络处理前向和后向的顺序数据,一个处理输入序列的正向部分,而另一个处理其反向部分。两个层的输出在每个时间步或序列结束时合并,增强了模型捕获双向上下文的能力。其结构如图 4 所示。

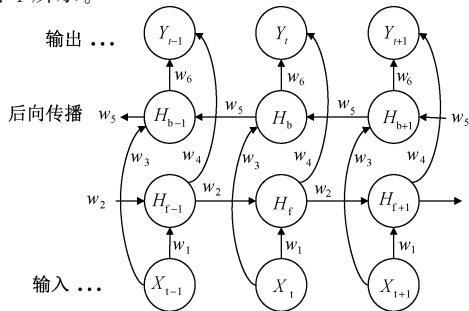


图 4 BiLSTM 结构图

Fig. 4 BiLSTM structure diagram

在前向传播层某一时刻 $t_i (i=1, 2, \dots, n)$ 正向计算, 将获得的时刻 t_i 前向隐含层的输出保存, 得到 H_f , 计算公式如式(10); 在后向传播层 $t_i (i=n, \dots, 2, 1)$ 反向计算, 将获得的时刻 t_i 后向隐含层的输出保存, 得到 H_b , H_b 计算公式如式(11); 然后将前向层和后向层每个时刻 t_i 获得对应的输出相结合获得最后的输出 Y_i , Y_i 计算公式如式(12), 其中 $w_i (i=1, 2, \dots, 6)$ 为权重。

$$H_f = f(w_1 x_t + w_2 H_f - 1) \quad (10)$$

$$H_b = f(w_3 x_t + w_5 H_b - 1) \quad (11)$$

$$Y_i = g(w_4 H_f + w_6 H_b) \quad (12)$$

研究中, 设计了新的 BiLSTM 网络结构, 通过残差连接通过非线性变换对输入进行处理, 并将处理后的结果与原始输入进行线性叠加输出。这种跳跃连接方式能够快速传递信息, 避免信息丢失和损耗, 从而解决了网络退化和梯度消失问题。可以利用非线性变换函数来描述神经网络的输入输出关系, 以第一次的残差连接为例, 输入为 X , 输出为 $F(X) + X$, F 在本模型中为 BiLSTM, 要拟合的函数可以表示为 $Y = H(x)$ 。引入残差连接之后, 拟合的目标可以表示为 $H(x) = X + F(X)$, 即 $F(X) = H(x) - x$ 。残差网络的学习目标由学习 $H(x)$ 变为学习残差 $F(x)$, 这样网络对于输出变化的微小扰动更加敏感, 改进的 BiLSTM 模块如图 5 所示。

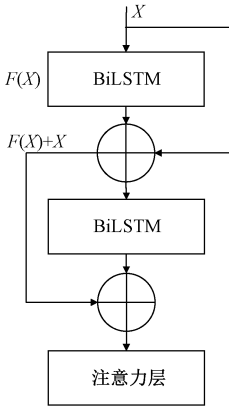


图 5 加入残差连接和注意力机制图

Fig. 5 Add residual connections and attention mechanism diagram

1.4 基于传感器数据时间信息的位置编码

传感器时间序列数据与机器翻译任务中的文本数据不同, 它们具有动态的时间特性而不是固定的语义关系。因此, 传统的位置编码方法无法有效地捕获传感器数据固有的时间信息。为了克服这一限制, 提出了一种改进的位置编码方法, 旨在动态地捕获传感器数据的时间动态。模型首先将输入序列分为查询部分和键值对部分, 然后将值部分与位置编码相结合。这里的位置编码是通过编码函数使用序列长度和输入特征维度生成, 以获得具有时间信息的增强表示。具体来说, 创建一个时间编码向量, 其范围从 0~1, 然后将时间信息映射到 $[0, 1]$ 的范围内, 用于表

示时间的相对位置。接着, 将时间编码向量扩展为与输入数据批次大小相匹配的二维向量, 并将其与传感器数据相加, 从而生成一个与输入数据对应的时间编码向量, 使得模型能够在处理序列数据时考虑到时间信息, 从而更好地理解序列中不同时间步长的相对位置关系, 改进的注意力机制如图 6 所示。

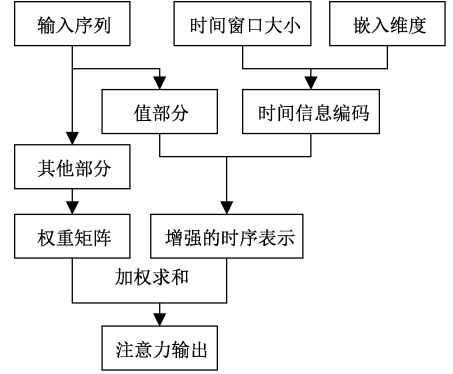


图 6 改进的注意力机制图

Fig. 6 Improved attention mechanism diagram

2 模型验证

2.1 模型验证数据库

本文采用了本田欧洲研究所采集的城市环境中自然日常行走的多模态步态数据库^[14]进行实验。该数据库包含了 20 名健康参与者 (其中 5 名女性和 15 名男性), 年龄介于 18~69 岁之间。在实验过程中, 参与者穿戴了带有多个 IMU 传感器的全身套装, 以及双脚脚底各有 8 个压力传感单元的鞋垫, 所有传感器的数据均已下采样至 60 Hz, 总共包含了 846 715 个采样数据点。本文选取了双足 16 个压力传感器和身体多部位三轴加速度传感器, 选取的三轴加速度和足底压力传感器的采集部位如图 7 所示。

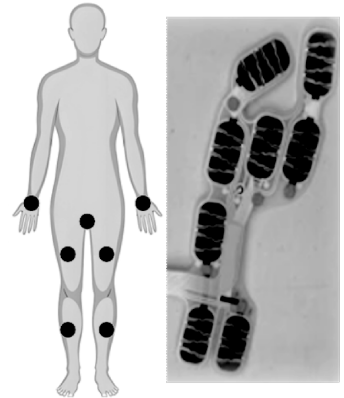


图 7 三轴加速度传感器与足底压力传感器分布图

Fig. 7 Distribution diagram of three-axis acceleration sensor and plantar pressure sensor

为了将原始的动作类型标签转换为模型可处理的数字标签,进行了标签映射操作,将动作类型标签‘走路’映射为数字 0,‘上楼梯’映射为数字 1,‘下楼梯’映射为数字 2,‘上坡’映射为数字 3,‘下坡’映射为数字 4。

2.2 数据预处理

首先对原始的.csv 类型文件进行预处理,将其提取并存储到.pickle 类型文件中,然后加载成为.pandas 类型数据。加速度传感器会因为磁场干扰和设备抖动等原因采集许多无关信号,同时,足底压力的采集过程中也会受到鞋底缓冲和摩擦等因素的影响,从而产生误差。为了最大程度地减少这些噪声对模型的干扰,需对原始信号进行平滑滤波处理。常见的数据平滑滤波方法包括高斯滤波、Savitzky-Golay 滤波和均值滤波等。其中,Savitzky-Golay 滤波器是一种特殊的低通滤波器,其主要功能是平滑噪声数据。图 8 显示了左大腿 x 轴加速度经 Savitzky-Golay 滤波前后的变化曲线,从图 8 可以观察到,Savitzky-Golay 滤波器能够在一定程度上平滑原始信号,同时保持边缘特征的清晰度,具有较高的边缘保持能力。通过对数据进行滤波预处理,能有效地消除噪声干扰,提高步态识别的准确性和稳定性。

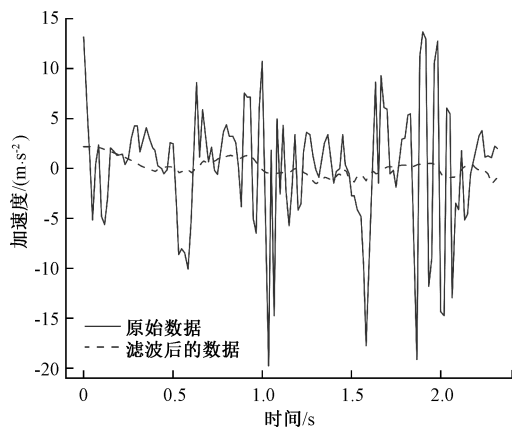


图 8 滤波前后示意图

Fig. 8 Schematic diagram before and after filtering

经足底压力传感器和 IMU 采集的数据是按照时间分布的数据流,很难直接对整个数据序列进行特征提取。因此,通常情况下采用的方法是利用滑动窗口技术对时间序列进行子序列提取,以确保提取出相同长度的子序列用于模型训练。原理是采用时间长度为 L 的窗口对数据进行分割,并以固定的步长 d 将窗口在时间轴上向后滑动,通常步长是窗口大小的一半。每个窗口对应的标签通常由窗口内占比最多的数据点对应的标签确定。在窗口包含多种标签时,这种方法会导致模型学习到片面的训练数据,影响模型的泛化能力。本模型在划分窗口时首先判断窗口内的数据点是否只对应一个标签。如果是,则保留该窗口及其对应的标签;否则丢弃该窗口。由于相邻窗口有重叠部分,若设置的步长足够小,则本次丢弃的数据已经

包含在前面的窗口中或者将被包含在下一个窗口中,不必担心数据丢失的问题。综合考虑,时间窗口大小设置为 1 s,步长设置为 0.33 s,图 9 为加速度数据的滑动窗口分割示意图。

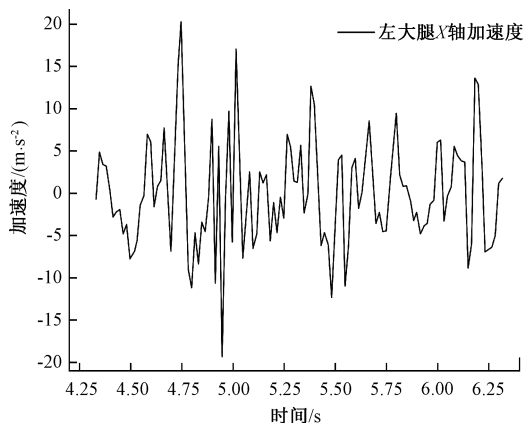


图 9 时间窗口划分图

Fig. 9 Time window division diagram

2.3 训练与结果分析

实验环境方面,本文提出算法使用的处理器为 NVIDIA GeForce RTX 4070 GPU,在 Python3.10 的环境下使用 TensorFlow 框架作为基础,保证实验的运行。为了有效控制模型的训练过程,采用了早停函数来避免过拟合。设置了早停函数监视验证集损失值,在连续 10 个 epochs 中损失值不再减小时中止训练。所有的实验数据均采用相同的输入数据。模型的训练过程中,设置数据的训练批量是 64,优化器为 Adam,学习率为 1×10^{-3} ;使用比例为 8:1:1 的训练集和验证集和测试集。实验结果均在测试集中得出。

基于多元时间序列的分类模型性能评估通常使用:准确率(Accuracy)、误差值(loss)、混淆矩阵等指标。准确率 A 表示模型正确预测的样本数与总样本数之比,计算公式为:

$$A = (TP + TN) / (TP + TN + FP + FN) \times 100\%$$

(13)

其中, TN 表示将负类预测为负类的数量, FN 表示将正类预测为负类的数量(漏报), TP 表示将正类预测为正类的数量, FP 表示将负类预测为正类的数量(误报)。混淆矩阵则是以表格形式展示了模型在不同类别上的预测结果与实际情况之间的对应关系。

为验证提出方法在多人活动识别的适用性和鲁棒性,在训练集和验证集上训练过程中的准确率和损失变化情况分别如图 10、11 所示。随着迭代次数的增加,训练集和验证集的准确率不断上升,损失逐步下降,模型拟合情况较好,最终模型在训练集上的准确率达到 99.48%,在验证集上的准确率达到 98.33%,混淆矩阵如图 12 所示。

为验证所提模型性能的优越性,将本文所提出模型与其他方法在测试集上进行对比实验,指标包括整体准确

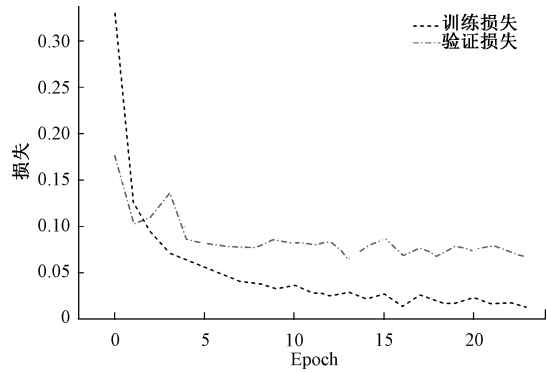


图 10 基于测试集的损失曲线图
Fig. 10 Loss curve graph based on test set

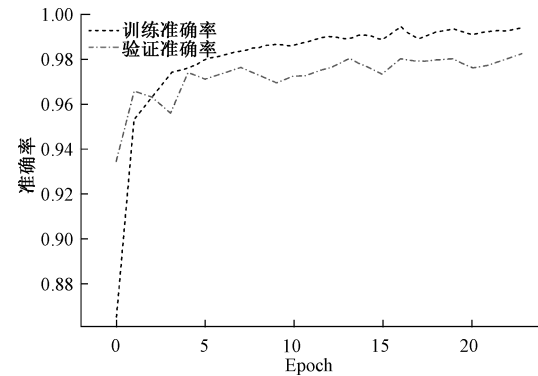


图 11 基于测试集的准确率曲线图
Fig. 11 Accuracy curve based on test set

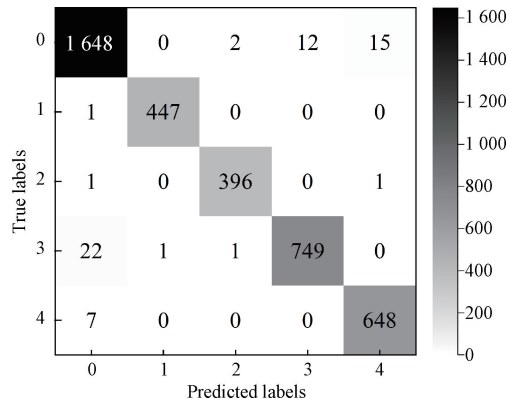


图 12 基于测试集的混淆矩阵图

Fig. 12 Confusion matrix diagram based on test set

率、整体精确率、整体召回率和平均单人准确率,结果如表 1 所示。

表 1 验证结果表明,模型 1 在整体准确率、精确率和召回率上表现优异,准确率达到 98.38%,显示出了很好的泛化能力和稳定性。模型 2 和模型 3 相较于模型 1 略有性能下降,但仍优于其他基线模型,表明 Transformer 特征提取模块和改进的 BiLSTM 模块在模型性能中扮演重要角色。模型 4 在去除改进的注意力模块后与模型 1 在整体指

表 1 不同方法在测试集上的指标

Table 1 Indicators of different methods on the test set

方法	整体 准确率	整体 精确率	整体 召回率	平均单人 准确率
模型	98.38	98.39	98.38	98.20
CNN-LSTM	97.32	98.37	97.21	97.77
C4M4BL	97.04	97.74	97.29	95.07
CNN	95.27	95.85	96.09	93.79

标上保持接近,表明改进的注意力模块对性能提升影响不大,但仍有助于整体性能。基于模型 1 的方法在整体表现上优于其他算法,包括 CNN-LSTM^[15]、C4M4BL^[16] 和 CNN^[17],进一步验证了所提出模型在人体行为识别任务上的优越性和有效性。因此,所提出的模型在人体行为识别领域具有显著的性能优势,为该领域的研究和应用提供了有力的支持和启示。

2.4 消融实验

在消融实验中设计并测试了五种不同的模型结构,以评估它们在时间序列分类任务中的性能。这些模型分别是 BiLSTM 模型(仅包含双向长短期记忆网络层)、BiLSTM+Attention 模型(在 BiLSTM 层后添加注意力机制层)、Attention+Transformer 模型(使用注意力机制和 Transformer 编码器)、BiLSTM+Transformer 模型(结合 BiLSTM 层和 Transformer 编码器)、以及 BiLSTM+Transformer+Attention 模型(在 BiLSTM 和变压器编码器后添加注意力机制层)。实验结果如表 2 所示。

表 2 不同模块融合在数据集上的准确度

Table 2 The accuracy of integrating different modules

方法	准确度/%
Bilstm	97.9
Bilstm+Attention	98.0
Attention+Transformer	97.5
Bilstm+Transformer	98.2
Bilstm+Transformer+Attention	98.4

实验结果显示,基础 BiLSTM 模型达到了 97.9% 的准确度,表现相当出色。这表明 BiLSTM 能够有效捕捉时间序列数据中的双向依赖关系,是处理此类任务的强大工具。在基础 BiLSTM 模型的基础上增加注意力机制后,BiLSTM+Attention 模型的准确度略微提升至 98.0%。注意力机制通过聚焦于输入序列中的关键部分,提高了特征提取的有效性,从而提升了模型性能。

Attention+Transformer 模型的准确度为 97.5%,略低于 BiLSTM 模型。这可能是由于 Transformer 需要更

大的数据量和更长的训练时间才能充分发挥其优势,或者是在特定任务上,LSTM 对局部时序模式的捕捉能力更强。结合了 BiLSTM 和 Transformer 编码器的 BiLSTM+Transformer 模型准确度提升至 98.2%。这种组合既能捕捉局部时序模式,又能建模全局依赖关系,从而提高了分类性能。

在 BiLSTM 和 Transformer 编码器后添加注意力机制的 BiLSTM+Transformer+Attention 模型达到了最高的 98.4% 的准确度。该模型结合了 BiLSTM、Transformer 和注意力机制的优势,进一步优化了特征提取和序列建模过程,是所有实验模型中表现最好的。

3 结 论

本研究提出了一种针对高维度、长时间序列分类的改进 Transformer-BiLSTM 网络模型。该模型采用多头注意力机制,能够并行高效处理数据,并重点关注不同维度之间的联系。同时,通过添加残差连接的双向 LSTM 层,更好地捕捉序列中的长期依赖关系,两次残差连接有效保留了关键信息。最后,设计了带有时间信息编码的注意力机制,赋予模型更灵活的序列数据理解能力。该方法通过将时间信息映射到固定范围内的连续空间,并将其与输入数据相结合,为模型提供了关于时间的相对位置信息。

在模型验证实验中,将所提出的网络模块与其他 3 种模型进行对比,并在公开数据集上进行了验证。结果显示,所提出的方法在测试集上达到了 98.38% 的准确率,明显高于其他深度学习方法,进一步提升了多人人体活动识别的准确率。这表明本研究提出的网络模型在处理高维度、长时间序列分类任务上具有优异的性能,为人体活动识别领域的研究和应用提供了重要的技术支持。

参考文献

- [1] YANG Y L, WANG H C, JIANG R ZH, et al. A review of IoT-enabled mobile healthcare: Technologies, challenges, and future trends[J]. IEEE Internet of Things Journal, 2022,9(12):9478-9502.
- [2] XU J, LI H, HOU SH J. Attention-based gait recognition network with novel partial representation PGOFI based on prior motion information[J]. Digital Signal Processing, 2023, 133:103845.
- [3] SU B B, SMITH C, GUTIERREZ F. Gait phase recognition using deep convolutional neural network with inertial measurement units[J]. Biosensors, 2020, 10(9):109.
- [4] HASSAN M M, UDDIN Z M, MOHAMED A, et al. A robust human activity recognition system using smartphone sensors and deep learning [J]. Future Generation Computer Systems, 2018, 81:307-313.

- [5] 吴海涛,陆志平,胡晨骏. 混合神经网络模型在人体活动识别中的研究[J]. 计算机应用与软件, 2021, 38(8):8.
WU H T, LU ZH P, HU CH J. Research on hybrid neural network model in human activity recognition[J]. Computer Applications and Software, 2021, 38(8): 8.
- [6] 王晓玲. 基于机器学习的人体活动识别[D]. 柳州:广西科技大学, 2023.
WANG X L. Human activity recognition based on machine learning[D]. Liuzhou: Guangxi University of Science and Technology, 2023.
- [7] PRAMANNIK R, SIKDAR R, SSARKAR R. Transformer-based deep reverse attention network for multi-sensory human activity recognition [J]. Engineering Applications of Artificial Intelligence, 2023, 122:106150.
- [8] MILON M I, SHEIKH N, FAKHRI K, et al. Multi-level feature fusion for multimodal human activity recognition in internet of healthcare things [J]. Information Fusion, 2023, 94: 17-31.
- [9] 魏淇,潘晴,田妮莉. 基于局部特征的时间注意力图卷积步态识别方法[J]. 国外电子测量技术, 2023, 42(10):19-24.
WEI Q, PAN Q, TIAN N L. Gait recognition method based on time attention graph convolution using local features [J]. Foreign Electronic Measurement Technology, 2023, 42(10): 19-24.
- [10] 陈万志,唐浩博,王天元. 融合轮廓增强和注意力机制的改进 GaitSet 步态识别方法[J]. 电子测量与仪器学报, 2024, 38(1):203-210.
CHEN W ZH, TANG H B, WANG T Y. Improved GaitSet gait recognition method integrating contour enhancement and attention mechanism[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 203-210.
- [11] SULLIVAN P O, MENOLOTTO M, VISENTIN A, et al. AI-based task classification with pressure insoles for occupational safety[J]. IEEE Access, 2024, 12: 21347-21357.
- [12] 张红颖,田鹏华. 结合残差网络与多级分块结构的步态识别方法[J]. 电子测量与仪器学报, 2022, 36(6):66-72.
ZHANG H Y, TIAN P H. Gait recognition method combining residual network and multi-level block structure[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(6): 66-72.
- [13] 邓帆,曾渊,刘博文,等. 基于 Transformer 时间特征聚合的步态识别模型[J]. 计算机应用, 2023, 43(S1): 15-18.

DENG F, ZENG Y, LIU B W, et al. Gait recognition model based on Transformer time feature aggregation[J]. Computer Applications, 2023, 43(S1): 15-18.

[14] LOSING V, HASENJÄGER M. A multi-modal gait database of natural everyday-walk in an urban environment[J]. Scientific Data, 2022, 9(1): 473.

[15] MEKRUKSAVANICH S, JITPATTANAKU A. Smart-watch-based human activity recognition using hybrid LSTM network [C]. 2020 IEEE Sensors, 2020: 1-4.

[16] SHAN C Y, HAN P Y, YIN O S. Deep analysis for smartphone-based human activity recognition [C]. 2020 8th International Conference on Information and Communication Technology(ICoICT), 2020: 1-5.

[17] 颜兵兵, 王强, 宋佳宝, 等. 基于足底压力分布的下肢步态识别方法[J]. 传感器与微系统, 2024, 43(2): 143-147.

YAN B B, WANG Q, SONG J B, et al. Lower limb gait recognition method based on plantar pressure distribution [J]. Transducer and Microsystem Technologies, 2024, 43(2): 143-147.

作者简介

孙巍伟, 副教授, 博士, 主要研究方向为机电设备状态监测和机器人技术。

E-mail: sww@bistu. edu. cn

毛亦鹏(通信作者), 硕士研究生, 主要研究方向为时间序列分类、步态识别。

E-mail: 1783518154@qq. com