

AHNNet:融合注意力机制的行为识别 混合神经网络模型

曹仰杰^{1,2}, 李昊¹, 段鹏松², 王福超², 王超²

(1. 郑州大学信息工程学院, 450000, 郑州; 2. 郑州大学软件学院, 450000, 郑州)

摘要: 针对 Wi-Fi 信号的行为感知研究中传统机器学习方法特征提取困难、深度学习方法特征提取方式单一, 导致特征提取不充分、识别准确率不高等问题, 提出融合注意力机制的人体行为识别混合神经网络模型 AHNNet。在对信道状态信息影响因子分析的基础上, 使用信道状态信息的振幅数据作为行为识别的基础数据; 采用时间滑窗将长时间人体活动序列分割为短时间序列, 构建样本数据, 克服全局人体行为数据非实时、长度不固定的缺点; 通过双向循环门控网络和时序卷积网络并行提取输入数据特征, 充分挖掘数据潜在特征之间的关系; 在双向循环门控网络中融合注意力机制以强化数据特征, 进一步提高模型性能; 将双向循环门控网络和时序卷积网络提取到的特征进行融合, 增加特征的多样性; 将融合特征输入到 Softmax 分类器进行分类, 得到人体活动数据对应的行为。与长短期记忆网络、双向循环神经网络进行了对比, 实验结果表明: 在标准数据采集室数据集上, AHNNet 的行为识别正确率达到 97.15%, 比未使用注意力机制的模型分类正确率提高 1.81%; 在公共数据集上, AHNNet 的行为识别正确率比其他对比模型的提高至少 0.65%, 参数量下降 47%; 在不同环境下, AHNNet 在卧室环境中的正确率为 95.7%, 比标准数据采集室中的下降 1.45%。AHNNet 具有良好的识别效果和鲁棒性, 并且在复杂的居家环境中应具有应用价值。

关键词: 人体行为识别; 信道状态信息; 双向循环门控网络; 时序卷积网络; 注意力机制

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202105014 **文章编号:** 0253-987X(2021)05-0123-10



OSID 码

AHNNet: Human Activity Recognition Based on Hybrid Neural Network Combining Attention Mechanism

CAO Yangjie^{1,2}, LI Hao¹, DUAN Pengsong², WANG Fuchao², WANG Chao²

(1. School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China;

2. School of Software, Zhengzhou University, Zhengzhou 450000, China)

Abstract: Activity wireless sensing is an important technology to implement health monitoring. Although the research on activity perception based on Wi-Fi has made good progress, there are still some problems, such as the difficulty of feature extraction in the traditional machine learning and the single feature extraction method in deep learning, which leads to insufficient feature extraction and low recognition accuracy. Here AHNNet, a hybrid neural network of human activity recognition integrating attention mechanism is proposed. Analyzing the influence factors of channel state information, the amplitude data of channel state information are used as the basic data for activity recognition, and the time sliding window is used to divide the long time human

activity sequence into short time series to construct the sample data, which solves the difficulties of non-real-time and non-fixed length global human activity data. AHNNet parallelizes the bidirectional gated recurrent network and the temporal convolutional network to extract the features of the input data, so as to fully reveal the relationship between potential features of data. To further improve the performance of model recognition, AHNNet is combined with attention mechanism to strengthen the main features of data in bidirectional gated recurrent network. Then, the features extracted from the bidirectional gated recurrent network and the temporal convolutional network are fused to increase the diversity of the features, and the fused features are input into the Softmax classifier for classification to obtain the activity corresponding to human activity data. Experimental results demonstrate that AHNNet has stronger classification ability than the other models, and the average accuracy rate achieves 97.15%, and AHNNet has fewer parameters while maintaining high accuracy. The competing model demonstrates that the accuracy rate is 95.7% in the bedroom environment, 1.45% lower than that in the standard data acquisition room. AHNNet has good recognition effect and robustness, especially in complex home environment.

Keywords: human activity recognition; channel state information; bidirectional gated recurrent network; temporal convolutional network; attention mechanism

随着人口老龄化进程加快,保障老年人特别是独居老人的健康生活,并对他们的日常行为进行监测是社会和家庭的一项长期义务^[1]。目前,已有不少算法可以实现跌倒^[2]和某些疾病的检测识别,如帕金森^[3]等,为医疗机构提供了及时救治时机和精准救治依据。为实现人体行为的有效检测识别,研究者们通常使用行为感知技术。行为感知技术是通过硬件或软件获取目标当前位置、行为和运动轨迹等信息,并对该信息进行分析与理解,实现对目标行为的感知。根据感知信息载体的不同,行为感知技术可以分为基于计算机视觉、基于专用传感器以及基于无线信号。

基于计算机视觉的感知技术是通过采集人体行为的图像序列或视频信息,利用计算机图形图像处理技术,从中提取出人体行为序列。常见应用有行为识别^[4]、跌倒检测^[5]、手势识别^[6]、步态识别^[7]等。该技术识别准确率高、应用范围广,缺点是易受光照和障碍物的影响,并且存在侵犯用户隐私和有监测死角等问题。

基于专用传感器的感知技术使用专用传感器或可穿戴设备采集相关人体行为,从而实现人体行为感知。文献[8]使用专用传感器来检测跑步、走路、转圈等多种人体行为。该技术可以实现细粒度的行为感知,准确率较高,但专用传感器安装和维护成本高,难以广泛应用。

基于计算机视觉和基于专用传感器的感知技术

虽然识别精度高,但存在侵犯个人隐私及需随身携带等缺点,限制了应用范围。基于无线信号的感知技术可以克服这些不足,近年来得到了研究者的广泛关注。基于无线信号的感知技术可简单分为基于特制射频信号和基于 Wi-Fi 信号两种。基于特制射频信号的感知技术需要特殊定制的设备,安装维护成本较高,不利于大规模使用^[9-10]。近年来,随着 Wi-Fi 热点的广泛部署^[11],利用 Wi-Fi 信号实现人体行为感知受到研究者的广泛关注。2011 年,Halperin 等发布了 CSI Tool 工具,极大方便了从商用 Wi-Fi 设备中提取信道状态信息(CSI)^[12]。CSI 包含丰富的振幅、相位信息,可以应用在细粒度的人体行为识别系统中,如睡眠检测^[13]、跌倒检测^[2]、手势识别^[14]、步态识别^[15-16]和行为识别^[17-19]等。

Wi-Fi 信号在传播过程中受多径效应影响,导致采集数据中包含环境和其他噪声,因此原始 CSI 数据不足以表征不同的人体行为。在实际应用中,传统机器学习方法,如文献[18]中提到的随机森林、隐马尔科夫模型,从原始 CSI 数据中人工提取相关行为特征,以区分不同人体行为。但是,这些方法不能充分挖掘时序数据中隐藏特征之间的关系,并且容易丢失特征,从而导致识别准确率不高。

随着深度学习技术的快速发展,大量深度学习模型用于人体行为感知技术中,以自动提取特征并提高识别性能^[17-19]。文献[18]利用长短期记忆网络(LSTM)自动提取 CSI 数据中的人体行为特征并进

行自主学习,该方法结果优于传统机器学习方法的。文献[19]采用双向循环神经网络 ABLSTM 从 CSI 数据中提取人体行为特征进行动作识别。文献[17]提出 WiSDAR 方法,将 CSI 数据中提取到的特征处理成图像帧,输入到二维卷积网络(CNN-2D)和 LSTM 拼接后的网络中,以实现人体行为识别。

LSTM 和 ABLSTM 使用循环神经网络(RNN)结构,以单一时间顺序方式提取人体行为数据特征,不能充分挖掘数据潜在特征之间的关系,并且模型的复杂度较高。虽然 WiSDAR 通过融合 CNN-2D 和 LSTM 使得网络结构更有效,但需要指定并按照前后顺序拼接两个独立的模型,而每个独立模型都有各自的复杂性,会阻止提取更细微的长期时空关系^[20]。

为解决传统机器学习方法特征提取困难、深度学习方法特征提取方式单一,导致潜在特征关系提取不充分、识别准确率不高等问题,本文设计了融合注意力机制的混合神经网络模型 AHNNet。该模型通过融合双向循环门控网络(BiGRU)和时序卷积网络(TCN)充分提取数据潜在特征。在 BiGRU 中,本文设计了引入注意力机制的双向循环门控网络 ABiGRU,以强化数据特征。实验结果表明,AHNNet 在本文采集的数据集上平均准确率较高,并且在保持高精度同时参数量更少。

1 人体行为识别模型

基于 AHNNet 的人体行为识别模型如图 1 所示,其主要包括数据处理模块和网络模块。在数据

处理模块,使用单发射端-单接收端构建数据采集平台,获取包含不同人体行为特征的 CSI 数据,并对采集到的数据使用滑动窗口,以提取更丰富的特征数据。网络模块实现融合注意力机制的混合神经网络模型 AHNNet,通过并行的两种不同特征提取方式的网络,充分挖掘 CSI 数据中的人体行为特征,提高模型准确率。之后,使用分类器对融合后的特征进行分类,实现不同人体行为的识别。

1.1 信道状态信息

CSI 是正交频分复用(OFDM)技术中估计通信链路信道属性的信息^[21]。如果将物理空间(包括环境对象和人)描述为无线信道,则信号在无线信道中传播会发生折射、衍射及散射现象。CSI 数据综合了信号在无线信道中传播的时间延迟、幅度衰减及相移的所有影响。在频域中,具有多个发射和接收天线的无线信道被描述为

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\theta}$$

(1)

式中 $\mathbf{y}$ 、 $\mathbf{x}$ 、 $\boldsymbol{\theta}$ 、 $\mathbf{H}$ 分别表示接收向量、发送向量、噪声向量以及信道矩阵。信道矩阵 $\mathbf{H}$ 是对 CSI 的估计,当环境中有人体移动时,信道矩阵会改变。因此,通过分析信道矩阵的变化,可以实现人体行为检测。

在 OFDM 技术中,CSI 以子载波的形式呈现,其中单个子载波可表示为

$$d = |d| \exp(j\sin\alpha)$$

(2)

式中: $|d|$ 表示振幅; α 表示相位。可以看出,CSI 能刻画子载波级别的信号扰动,提供了细粒度的无线信道描述。

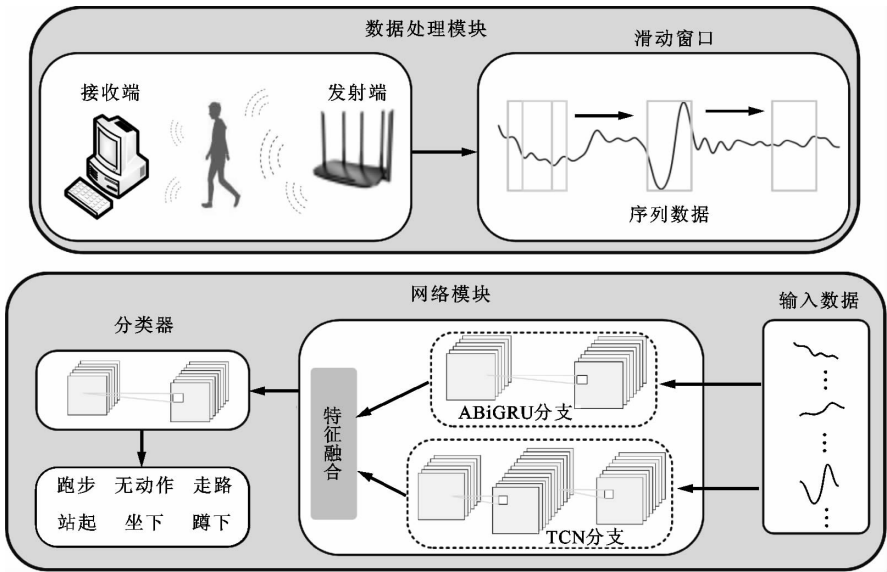


图 1 人体行为识别模型

Fig. 1 Human activity recognition model

1.2 时间滑窗

数据处理模块中使用时间滑动窗口,有助于提取更加丰富的特征信息,并丢弃标记为 NoActivity (无动作)的冗余数据。在滑动窗口内,如果超过40%的采样数据被标记为 NoActivity,则该滑窗内的所有数据将被丢弃。随着滑动窗口后移,标记为有效活动的数据被不断地选择用于训练或测试。图2是时间滑窗及滑动步长示意。

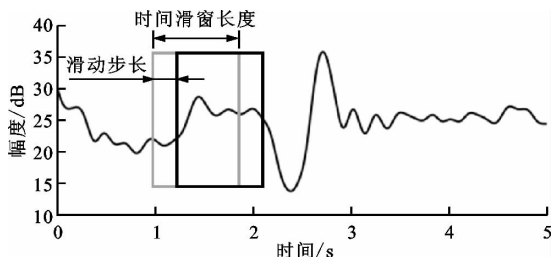


图2 时间滑窗及滑动步长示意

Fig. 2 Schematic of time sliding window and sliding step

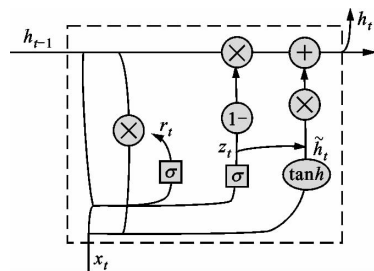
滑动窗口将长时间人体活动序列数据分割成短时间序列,然后从中提取数据特征。该处理方式可以关联活动序列的上下文信息,克服了全局人体行为数据非实时、数据量固定的缺点。

1.3 AHNNet 模型

AHNNet 采用的主要技术包括 BiGRU、TCN 及注意力机制等,本小节将详细介绍这些技术及模型构建细节。

1.3.1 双向循环门控网络 传统 RNN 及其变体(如长短期记忆网络^[22]、LSTM)在提取时序数据特征时,只能记忆过去的信息,而对于人体行为数据,未来信息对当前时刻也有重要意义。因此,本文在 AHNNet 中采用以 GRU^[23] 为基本神经元的 BiGRU^[24],用于同时提取人体行为数据过去和未来的信息,提高模型识别性能。GRU 是 LSTM 的一种变体,其内部只有两个信息控制单元(更新门和重置门),而 LSTM 有 3 个信息控制单元(输入门、遗忘门和输出门)。因此,GRU 减少了网络单元内部计算复杂度,从而加快了网络训练和收敛速度。另外,本文中的行为数据经过滑动窗口处理后变得不那么“长”,亦比较适用 GRU 的运行。GRU 的内部结构如图 3 所示。

更新门 z_t 用于控制前一时刻 t 的状态信息被带入到当前时刻 t 的状态中的程度。更新门的数值越大,说明 $t-1$ 时刻状态信息带入越多。重置门 r_t 控制 $t-1$ 时刻有多少状态信息被写入到当前时刻 t 的候选集 $\tilde{h}_t$ 上。重置门的数值越小, $t-1$ 时刻的状



x_t —输入; z_t —更新门; r_t —重置门; σ —激活函数;
 h_{t-1} —前一时刻状态信息; $\tilde{h}_t$ —候选信息; h_t —状态信息。

图3 GRU 的内部结构

Fig. 3 The internal structure of GRU

态信息被写入的越少。 h_t 表示传递到下一个节点的隐藏状态。

如果用 o_t 表示当前时刻 t 的 GRU 输出值, $\rightarrow$ 表示提取过去信息, $\leftarrow$ 表示提取未来信息, $\oplus$ 表示信息拼接,则 BiGRU 的输出为

$$o_t = \vec{o}_t \oplus \overleftarrow{o}_t \quad (3)$$

1.3.2 时序卷积网络 TCN 在时序数据特征提取上已经证明是有效的^[20]。行为数据也是时序数据的一种,因此本文尝试将 TCN 引入到 AHNNet 中。GRU 以时间顺序提取行为数据特征,而 TCN 在时间尺度上分层提取行为数据特征。本文将两者有效融合,能充分挖掘数据特征之间的关系,从而提高模型性能。

在 TCN 中,令 $\mathbf{X}_t \in \mathbf{R}^{F_0}$ 为 t 时刻 ($0 < t \leq T$) 长度为 F_0 的输入特征向量。值得注意的是,本文中时间 T 固定,取为 800,与滑动窗口长度一致。每个时间序列的真实标签由 $p_i \in \{1, \dots, c\}$ 给出,其中 c 是类别数。假设网络模型有 L 层卷积层,每层卷积使用一组一维卷积核提取输入的人体行为数据在时间维度上的变化特征。每层卷积核由权重张量 $\mathbf{W}^l \in \mathbf{R}^{F_l \times d \times F_{l-1}}$ 和偏置 $b^l \in \mathbf{R}^{F_l}$ 参数化,其中 l 是层索引, d 是卷积核尺寸。第 l 个卷积层的输出 $\mathbf{E}_{i,t}^l \in \mathbf{R}^{F_l}$ 是来自前一层输入矩阵 $\mathbf{E}_{i,t}^{l-1} \in \mathbf{R}^{F_{l-1} \times T}$ 的函数

$$\mathbf{E}_{i,t}^l = f_{\text{relu}} \left(b_i^l + \sum_{\tau=1}^d \langle \mathbf{W}_{i,\tau}^l, \mathbf{E}_{i,t+d-\tau}^{l-1} \rangle \right) \quad (4)$$

式中: $\langle \cdot \rangle$ 表示相关函数; $f_{\text{relu}}(\cdot)$ 表示激活函数 ReLU。

1.3.3 注意力机制 传统人体识别模型(如 LSTM 等)在提取人体行为数据特征时,认为所有特征同等重要。但是,人体行为数据不仅包含行为信息,还包含大量噪声(如墙壁、家具等静态物体的反射信号)。为使 AHNNet 更多关注于人体行为特征,本文将注意力机制^[25] 融合到网络。注意力机制

和人类视觉的注意力类似,就是在众多信息中把注意力集中放在重要的部分,选出关键信息而忽略其他不重要的信息。注意力机制最早应用在图像识别中,近年来已经被大量应用于时序数据处理中(如自然语言处理^[26]等)。本文将注意力机制应用到人体行为识别中,让模型更关注当前信息对当前动作的重要程度,从而提高模型识别性能。

注意力机制的核心是权重参数,首先学习每个元素的重要程度,然后按重要程度为每个元素分配权重,权重越大则重要程度越高。假设输入数据为 $\mathbf{b}_t = [b_1, b_2, \dots, b_n]$, $0 < t \leq T$, n 为序列长度。注意力机制首先提取输入数据 $\mathbf{b}_t$ 中的特征向量

$$\mathbf{C}_t = f(\mathbf{b}_t) \quad (5)$$

式中 $f(\cdot)$ 可以是 CNN、RNN 或全连接网络等。然后,计算 $\mathbf{C}_t$ 向量中每个特征所占权重向量

$$\mathbf{w}_{ti} = \frac{\exp(\mathbf{C}_{ti})}{\sum_{i=1}^n \exp(\mathbf{C}_{ti})} \quad (6)$$

最后,将输入数据 $\mathbf{b}_t$ 与其权重向量 $\mathbf{w}_t$ 相乘求和,得到最终输出结果

$$\mathbf{o} = \sum \mathbf{b}_t \cdot \mathbf{w}_t \quad (7)$$

1.3.4 分类器模块 在 AHNNet 中,人体行为类别由 $p_i \in \{1, 2, \dots, c\}$ 给出,根据输入数据 $\mathbf{x}$ 可以预测出可能的人体行为类别 p 。AHNNet 旨在解决多分类问题,因此使用 Softmax 函数更为合适,公式为

$$\bar{p}_i = P(p_i | \mathbf{x}) = \frac{\exp(\mathbf{z}_i)}{\sum_{i=1}^c \exp(\mathbf{z}_i)}, i \in [1, c] \quad (8)$$

式中: $\bar{p}_i$ 为每个输出类别的概率; $\mathbf{z}_i$ 为融合后的特征数据。得到 $\bar{p}_i$ 后,利用 Adam 优化器对预测标签和真实标签之间的交叉熵损失进行最小化训练,计算公式为

$$L = - \sum p_i \log(\bar{p}_i) \quad (9)$$

经过多次迭代,Adam 优化器可以找到较优模型参数,实现对人体行为的精准识别。

1.3.5 模型构建 图 4 是 AHNNet 模型结构,输入数据经过输入层后分别输入左分支 TCN 网络和右分支 ABiGRU 网络进行特征提取,然后将两分支提取到的特征进行融合,最后输入到 Softmax 分类器进行分类。

TCN 分支主要由 3 个 TCN 块组成,每个 TCN 块包含 1 个一维卷积层、1 个归一化层^[27]和 1 个激活层。其中:第 1 个 TCN 块中不添加 Dropout 层,

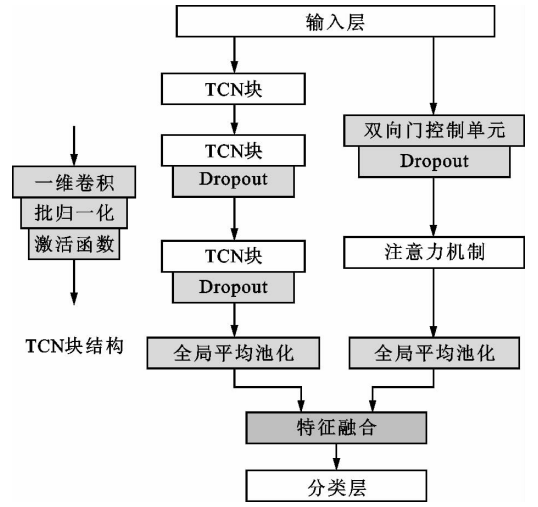


图 4 AHNNet 模型结构

Fig. 4 Structure of AHNNet model

以最大可能保留低级特征,防止特征丢失;第 2 个和第 3 个 TCN 块后添加 Dropout 层^[28],以防止模型过拟合。数据输入到一维卷积层进行特征提取后,为防止数据分布在训练过程中发生改变,AHNNet 使用批归一化对数据进行处理,然后将数据输入到激活层进行非线性变换,再输入到 Dropout 层,最后使用全局平均池化^[29],其优点是可以在对整个 TCN 网络分支进行正则化,防止过拟合且可以减少模型参数量。

ABiGRU 分支由两部分组成。使用 GRU 神经元为基础的 BiGRU 同时提取人体行为数据过去和未来的特征,并输入到 Dropout 层;将提取的特征应用注意力机制,为不同特征分配不同权重,以区分不同特征对当前动作的重要程度;使用全局平均池化调整数据维度,方便后面两个分支特征进行融合。

为实现注意力机制,先使用全连接层提取输入数据特征,并使用 Softmax 函数计算权重向量,最后将输入数据与其权重向量相乘求和得到输出结果。注意力机制结构如图 5 所示。

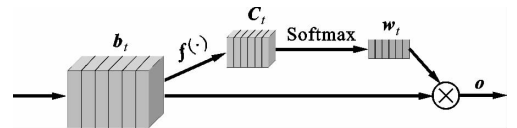


图 5 注意力机制结构

Fig. 5 Attention mechanism structure

得到 TCN 分支和 ABiGRU 分支的输出特征后,AHNNet 将二者融合,输入到 Softmax 分类器进行分类。

2 实验

2.1 实验设置

本文实验使用深度学习框架 Keras 实现。AHNNet 模型中 3 个 TCN 块所使用的卷积核数分别为 128、256 和 128, 长度分别为 8、5 和 3, ABiGRU 的神经元数为 128, Dropout rate 设置为 0.7。本文使用 Adam 优化器, 学习率的初始值设置为 $\eta = 0.001$, 动量参数设置为 $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ 。为加快模型训练和收敛速度, 将 batch_size 设置为 16。本文中数据集分为训练集、验证集和测试集, 比例分别为 0.8、0.1 和 0.1, 其中验证集用来调整模型超参数。

2.2 数据集描述

本文使用 5 个数据集评估 AHNNet 的性能, 其中一个公共数据集, 其余 4 个是本文在不同环境下采集自建的数据集。

公共数据集^[29]在室内办公区域采集。发射器和接收器在视距条件下相距 3 m。接收器是配备商用 Intel 5300 网卡的笔记本电脑。在数据采集过程中, 每个实验者的活动时间为 20 s, 并需要在视距条件下进行一项活动。在该时间段的开始和结束, 实验者保持静止。整个数据收集过程由摄像机记录, 以便对数据进行标记。该数据集包括 6 个人的 6 个活动, 分别为躺下、跌倒、走路、跑步、坐下和站起。公共数据集在使用时间滑窗去除无动作的时间片段后, 剩下有效(即包含动作)时长约 3 s。

为综合评估环境对 AHNNet 性能的影响, 本文另外采集了 5 个不同环境的数据集。采集环境分别是标准数据采集室、会议室、展厅、心理咨询室和卧室, 其布局如图 6 所示。5 种环境使用相同设备及配置进行数据采集: TP_LINK AC1750 路由器为发射器, 配备 Intel 5300 网卡的台式计算机为接收器。为使数据标记更准确, 在数据采集过程中, 要求实验者在一段时间内重复某项动作, 每次采集其中 5 s 的数据片段作为一个动作序列。为减少操作开始和结束对数据的影响, 将动作序列前后 1 s 视为无动作。为使数据更加多样化, 本文收集 10 名实验者的 6 种常见行为, 包括无动作、走路、慢跑、坐下、站起和蹲下。

所有数据集采样频率均为 1 kHz。接收器有 3 根天线, 每根接收 30 个子载波。数据集规模方面, 公共数据集经处理后有 3 400 条数据用于训练, 本文 5 个数据集处理后各有 3 600 条数据。经实验验证,

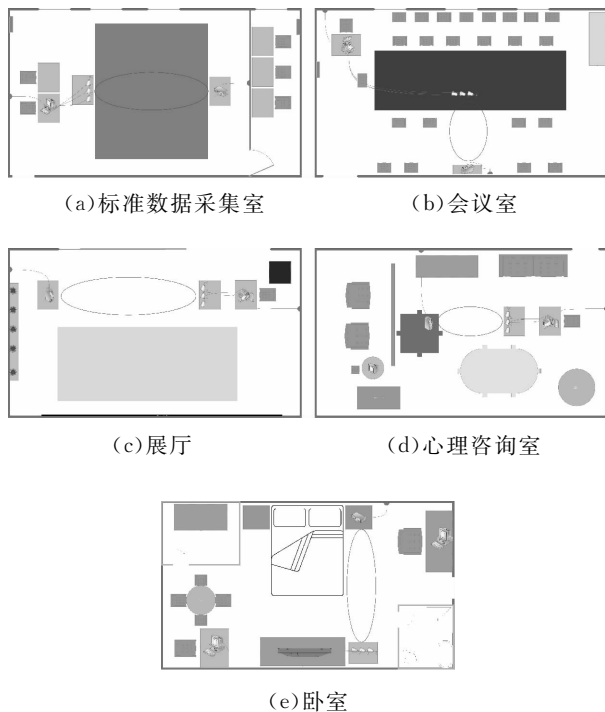


图6 不同实验环境布局

Fig. 6 Different experimental environment layout

本文所用数据集规模适合轻量级的 AHNNet 模型。

2.3 评价指标

分类模型的评价指标通常为准确率 A 、精准率 P 、召回率 R 、综合评价指标 F_1 分数。准确率 A 表示算法所有分类正确的样本占所有数据样本的比率, A 越高则算法检测数据样本真实情况的能力越强。精准率 P 和召回率 R 都表示在不同分类情况下对该类别的判断能力, P 和 R 越高, 则该算法对该类别的分类能力越强。为了避免 P 和 R 一个为 1 而另一个为 0 的极端情况, 使用 P 和 R 的调和平均值 F_1 来评估模型的性能。各项评价指标公式为

$$\left. \begin{aligned} A &= \frac{S_{TP} + S_{TN}}{S_{TP} + S_{TN} + S_{FN} + S_{FP}} \\ P &= \frac{S_{TP}}{S_{TP} + S_{FP}} \\ R &= \frac{S_{TP}}{S_{TP} + S_{FN}} \\ F_1 &= \frac{2PR}{P + R} \end{aligned} \right\} \quad (10)$$

式中: S_{TP} 为属于该类且被分类器分为该类的样本数; S_{FP} 为不属于该类但被分类为该类的样本数; S_{FN} 为属于该类但被分类为不属于该类的样本数; S_{TN} 为不属于该类且被分类为不属于该类的样本数。

2.4 实验结果

为综合评估 AHNNet 的性能, 本文从滑窗参数

设置、模型结构选择、不同数据集上模型性能、不同环境下模型适应能力共 4 个方面出发,设计了一系列实验进行验证。另外,还从参数量和浮点计算次数方面将 AHNNet 与同类模型的复杂度进行了对比分析。

2.4.1 时间滑窗长度及滑动步长选择 不同长度的时间滑窗对行为数据细化程度不同,窗口内包含的特征完整度不同,对模型性能有较大影响。为选择合适的时间滑窗长度,本文进行多组实验,结果如图 7 所示。可以看出,当时间滑窗长度 $T=800$ 时,AHNNet 模型性能最佳。因此,本文选择 $T=800$ 。

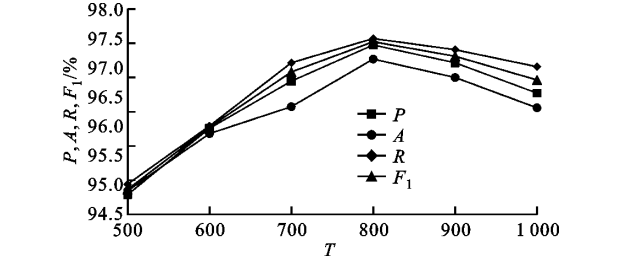


图 7 不同时间滑窗长度下的实验结果

Fig. 7 Experimental results for different sliding window lengths

此外,本文进行另一组实验来选择合适的滑动步长。选择滑动步长为 100、200、300、400、600、800 共 6 组数据进行实验。随着滑动步长的增加,每个 CSI 序列被分割成的子序列数量和子序列之间的重叠率将减少。不同滑动步长和重叠率下的模型准确率如表 1 所示,可以看出,当滑动步长为 200 时,模型性能比较好。因此,本文选择 200 作为滑动步长。

表 1 不同滑动步长和重叠率下的模型准确率						
Table 1 Accuracy for different sliding step and repetition rate						
滑动步长	100	200	300	400	600	800
重叠率	0.875	0.750	0.625	0.500	0.250	0
准确率	0.955	0.971	0.962	0.951	0.949	0.940

2.4.2 模型结构对比 模型结构对模型性能有重要影响。为验证注意力机制、Dropout 层和 GRU 的效果,设计了不添加注意力机制、不添加 Dropout 层、使用 LSTM 神经元和 AHNNet 这 4 种网络结构。在标准数据采集室采集的数据集上测试了 4 种主要的网络结构,结果如表 2 所示。可以看出:在不添加注意力机制的情况下,模型性能有所下降,说明注意力机制可以提升模型对主要特征的感知程度,提高模型性能;Dropout 层能够明显提升模型的泛化

性能;GRU 神经元的性能指标略微高于 LSTM 的。

表 2 不同网络结构的实验结果

Table 2 Experimental results of different network structures

模型名称	P	A	R	F ₁
AHNNet 不添加注意力机制	95.61	95.34	95.9	95.62
AHNNet 不添加 Dropout 层	96.20	95.85	96.37	96.11
AHNNet 使用 LSTM 神经元	96.98	96.63	97.08	96.85
AHNNet	97.39	97.15	97.50	97.32

图 8 是 GRU 和 LSTM 的损失函数,可以看出,在训练过程中 GRU 比 LSTM 更稳定且收敛更快。综合表 2 和图 8,本文选用 GRU 作为神经元。

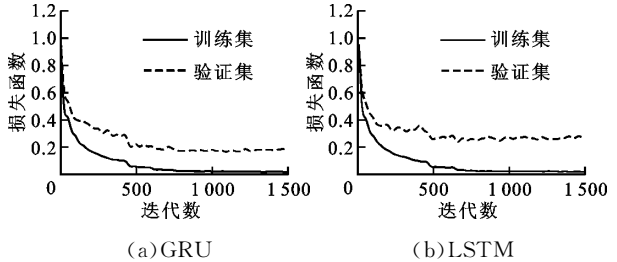


图 8 GRU 和 LSTM 的损失函数

Fig. 8 Loss function of GRU and LSTM

2.4.3 公共数据集上的对比 混淆矩阵可以直观呈现模型对每个类别的分类性能。本文以 LSTM^[18] 和 ABLSTM^[19] 为基准实验,验证在公共数据集上各模型的性能。LSTM^[18]、ABLSTM^[19] 和 AHNNet 在公共数据集上的混淆矩阵如图 9 所示。可以看出,AHNNet 不仅在整体分类性能上表现更好,而且在各个类别上也取得了相对较好性能。这主要由于 ABiGRU 和 TCN 分别提取数据特征再融合的机制,使得网络特征提取能力增强。另外,混淆矩阵显示每个类别分类错误都是在相似动作上,这是因为相似动作的特征相似度更高,更容易导致分类错误。

		预测的动作					
		躺下	跌倒	走路	跑步	站起	坐下
真实的动作	躺下	0.95	0.01	0.01	0.01	0	0.02
	跌倒	0.01	0.94	0.05	0	0	0
	走路	0	0.01	0.93	0.04	0.01	0.01
	跑步	0	0	0.02	0.97	0.01	0
	站起	0.03	0.01	0.05	0.02	0.81	0.07
	坐下	0.01	0	0.03	0.05	0.07	0.83

(a)LSTM

		预测的动作					
		躺下	跌倒	走路	跑步	站起	坐下
真实的动作	躺下	0.96	0	0.01	0	0.02	0.01
	跌倒	0	0.99	0	0.01	0	0
	走路	0	0	0.98	0.02	0	0
	跑步	0	0	0.02	0.98	0	0
	站起	0.01	0.01	0.01	0	0.95	0.02
	坐下	0.01	0	0	0	0.01	0.98

(b) ABLSTM

		预测的动作					
		躺下	跌倒	走路	跑步	站起	坐下
真实的动作	躺下	0.974	0	0	0.026	0	0
	跌倒	0	1	0	0	0	0
	走路	0	0	0.987	0.013	0	0
	跑步	0	0	0.018	0.982	0	0
	站起	0	0	0	0	0.97	0.03
	坐下	0	0	0	0	0.034	0.996

(c) AHNNet

图 9 LSTM^[18]、ABLSTM^[19] 和 AHNNet 在公共数据集上的混淆矩阵

Fig. 9 Confusion matrix of LSTM^[18], ABLSTM^[19] and AHNNet on common dataset

2.4.4 不同环境实验下的对比 为验证模型在不同环境下的适应性,本文在 5 个环境中的数据集使用 AHNNet 进行实验,结果如图 10 所示。可以看出:标准数据采集室的结果最好,这是因为标准数据采集室是一个理想的实验环境,障碍物最少,无关信号反射对人体行为影响最小;会议室和展厅的结果稍差,这是因为这两种环境中与实验无关的障碍物,但相对较少;卧室和心理咨询室的结果最差,这是因为这两种环境最为复杂,有许多非实验因素。综合分析可知,虽然 AHNNet 模型的性能会受障碍物的影响,但在不同实验环境中具有良好的适应性和鲁棒性。

2.4.5 复杂度分析 评估模型复杂度常用的指标是模型参数数量和浮点计算次数。LSTM、ABLSTM 及 AHNNet 的参数量分别为 0.802×10^6 、 2.04×10^6 、 1.08×10^6 ,浮点计算次数分别为 1.28×10^6 、 3.44×10^6 、 1.97×10^6 。可以看出,AHNNet 模型在复杂度上相对偏低,原因有两点:① LSTM 和

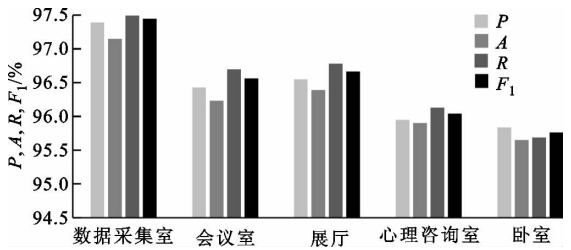


图 10 不同环境下的实验结果

Fig. 10 Experimental results in different environments

ABLSTM 的神经元数为 200,而 AHNNet 的 ABiGRU 分支神经元数为 128,并且使用参数量较少的 GRU,从整体上减少了模型参数量及计算量;②相比于 GRU 在计算时因保存每步信息而占据大量内存,TCN 网络层内卷积核是共享的,因此参数量和计算量更少,对内存的占用也更低。虽然 AHNNet 模型结合了两种不同的网络,但是其复杂度并没有大幅度增加,反而相对比较轻量化,同时拥有较高精度。

3 结 论

本文提出一种用于日常行为健康监测的融合注意力机制的混合神经网络模型 AHNNet。该模型将融合注意力机制的双向循环门控网络与时序卷积网络并行,以不同方式提取数据特征,并且在保持高识别精度的情况下能有效降低模型规模。在公共数据集和本文不同环境数据集上,AHNNet 均表现出良好的识别和分类性能。不同环境下的实验结果表明了 AHNNet 良好的鲁棒性。

未来可从以下 3 方面进行进一步研究:①多人行为检测;②CSI 的相位信息在人体行为识别中的有效性及使用方法;③提高模型的环境自适应能力。

参考文献:

[1] GHASEMZADEH H, JAFARI R. Physical movement monitoring using body sensor networks: a phonological approach to construct spatial decision trees [J]. IEEE Transactions on Industrial Informatics, 2011, 7(1): 66-77.

[2] WANG Yuxi, WU Kaishun, NI L M. WiFall: device-free fall detection by wireless networks [J]. IEEE Transactions on Mobile Computing, 2017, 16(2): 581-594.

[3] WANG Tianben, WANG Zhu, ZHANG Daqing, et al. Recognizing parkinsonian gait pattern by exploiting fine-grained movement function features [J]. ACM

- Transactions on Intelligent Systems and Technology, 2016, 8(1): 1-22.
- [4] SHI Lei, ZHANG Yifan, CHENG Jian, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA; IEEE, 2019: 12026-12035.
 - [5] NGUYEN H T K, FAHAMA H, BELLEUDY C, et al. Low power architecture exploration for standalone fall detection system based on computer vision [C]// Proceedings of the 2014 European Modelling Symposium. Piscataway, NJ, USA; IEEE, 2014: 169-173.
 - [6] JIANG Du, ZHENG Zujia, LI Gongfa, et al. Gesture recognition based on binocular vision [J]. Cluster Computing, 2019, 22(S6): 13261-13271.
 - [7] RAMAKIC A, BUNDALO Z, BUNDALO D. A method for human gait recognition from video streams using silhouette, height and step length [J]. Journal of Circuits, Systems and Computers, 2020, 29(7): 2050101.
 - [8] LARA Ó D, LABRADOR M A. A survey on human activity recognition using wearable sensors [J]. IEEE Communications Surveys and Tutorials, 2013, 15(3): 1192-1209.
 - [9] GU Yu, WANG Yantong, LIU Zhi, et al. Sleep-Guardian: an RF-based healthcare system guarding your sleep from afar [J]. IEEE Network, 2020, 34(2): 164-171.
 - [10] ZHAO Mingmin, LI Tianhong, ALSHEIKH M A, et al. Through-wall human pose estimation using radio signals [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA; IEEE, 2018: 7356-7365.
 - [11] Cisco Mobile. Cisco visual networking index: global mobile data traffic forecast update 2016-2021 [EB/OL]. [2020-09-01]. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf.
 - [12] HALPERIRR D, HU W, SHETH A, et al. Tool release: gathering 802.11n traces with channel state information [J]. Computer Communication Review, 2011, 41(1): 53-53.
 - [13] CAO Yangjie, WANG Fuchao, LU Xinxin, et al. Contactless body movement recognition during sleep via WiFi signals [J]. IEEE Internet of Things Journal, 2020, 7(3): 2028-2037.
 - [14] LI Hong, YANG Wei, WANG Jianxin, et al. WiFin-ger: talk to your smart devices with finger-grained gesture [C]// Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, USA; ACM, 2016: 250-261.
 - [15] 段鹏松, 周志一, 王超, 等. WiNet: 一种适用于无线感知场景的步态识别模型 [J]. 西安交通大学学报, 2020, 54(7): 187-195.
 - DUAN Pengsong, ZHOU Zhiyi, WANG Chao, et al. WiNet: a gait recognition model suitable for wireless sensing scene [J]. Journal of Xi'an Jiaotong University, 2020, 54(7): 187-195.
 - [16] 余星达, 陈文杰, 王鼎, 等. 非接触式身份识别的深度学习算法 [J]. 西安交通大学学报, 2019, 53(4): 122-127.
 - YU Xingda, CHEN Wenjie, WANG Ding, et al. A deep learning algorithm for contactless human identification [J]. Journal of Xi'an Jiaotong University, 2019, 53(4): 122-127.
 - [17] WANG Fangxin, GONG Wei, LIU Jiangchuan. On spatial diversity in WiFi-based human activity recognition: a deep learning-based approach [J]. IEEE Internet of Things Journal, 2019, 6(2): 2035-2047.
 - [18] YOUSEFI S, NARUI H, DAYAL S, et al. A survey of human activity recognition using wifi CSI [EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1708.07129.pdf>.
 - [19] CHEN Zhenghua, ZHANG Le, JIANG Chaoyang, et al. WiFi CSI based passive human activity recognition using attention based BLSTM [J]. IEEE Transactions on Mobile Computing, 2018, 18(11): 2714-2724.
 - [20] LEA C, VIDAL R, REITER A, et al. Temporal convolutional networks: a unified approach to action segmentation [M]// Lecture Notes in Computer Science. Cham, Germany; Springer, 2016: 47-54.
 - [21] WU K S, XIAO J, YI Y W, et al. FILA: fine-grained indoor localization [C]// Proceedings of the IEEE INFOCOM. Piscataway, NJ, USA; IEEE, 2012: 2210-2218.
 - [22] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
 - [23] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB/OL]. [2020-09-01]. <https://arxiv.org/abs/1412.3555>.
 - [24] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks [J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.

- [25] DENIL M, BAZZANI L, LAROCHELLE H, et al. Learning where to attend with deep architectures for image tracking [J]. *Neural Computation*, 2012, 24(8): 2151-2184.
- [26] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [EB/OL]. [2020-09-01]. <https://arxiv.org/abs/1409.0473>.
- [27] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. [2020-09-01]. <https://arxiv.org/abs/1502.03167>.
- [28] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. *Journal of Machine Learning Research*, 2014, 15: 1929-1958.
- [29] LIN Min, CHEN Qiang. Network in network [EB/OL]. [2020-09-01]. <https://arxiv.org/pdf/1312.4400.pdf>.

[本刊相关文献链接]

- 孙红帅,王霞,柳莹,等. 频域注意力机制下的癫痫脑电信号分类. 2021,55(2):129-135. [doi:10.7652/xjtuxb202102015]
- 谢雨杰,杜友田,张潇. 面向新概念学习的图像描述生成模型. 2020,54(12):37-44. [doi:10.7652/xjtuxb202012005]
- 张西宁,郭清林,刘书语. 深度学习技术及其故障诊断应用分析与展望. 2020,54(12):1-13. [doi:10.7652/xjtuxb202012001]
- 张连超,乔瑞萍,党祺玮,等. 具有全局特征的空间注意力机制. 2020,54(11):129-138. [doi:10.7652/xjtuxb202011016]
- 张克旭,杜昌旺,赵浩淇,等. 针对局灶性癫痫患者的脑电微状态分析. 2020,54(9):157-163. [doi:10.7652/xjtuxb202009018]
- 景海婷,张秦,陈曼,等. 小样本域自适应的皮肤病临床影像识别方法. 2020,54(9):142-148,156. [doi:10.7652/xjtuxb202009016]
- 赵小强,梁浩鹏. 使用改进残差神经网络的滚动轴承变工况故障诊断方法. 2020,54(9):23-31. [doi:10.7652/xjtuxb202009002]
- 张弘斌,袁奇,赵柄锡,等. 采用多通道样本和深度卷积神经网络的轴承故障诊断方法. 2020,54(8):58-66. [doi:10.7652/xjtuxb202008008]

段鹏松,周志一,王超,等. WiNet:一种适用于无线感知场景的步态识别模型. 2020,54(7):187-195. [doi:10.7652/xjtuxb202007022]

孟月波,纪拓,刘光辉,等. 编码-解码多尺度卷积神经网络人群计数方法. 2020,54(5):149-157. [doi:10.7652/xjtuxb202005020]

余艳杰,孙嘉琪,葛思擘,等. CycleGAN-SN:结合谱归一化和CycleGAN的图像风格化算法. 2020,54(5):133-141. [doi:10.7652/xjtuxb202005018]

张帆,张景森,裴昌幸. 采用双通道时间稠密网络的高光谱影像分类方法. 2020,54(4):126-132. [doi:10.7652/xjtuxb202004016]

刘天宇,周稻祥,李明,等. 面向螺丝锁附序列的多分辨率融合卷积神经网络. 2020,54(3):161-168,178. [doi:10.7652/xjtuxb202003020]

吴静然,丁恩杰,崔冉,等. 采用多尺度注意力机制的旋转机械故障诊断方法. 2020,54(2):51-58. [doi:10.7652/xjtuxb202002007]

毛远宏,贺占庄,马钟,等. 采用类内迁移学习的红外/可见光异源图像匹配. 2020,54(1):49-55. [doi:10.7652/xjtuxb202001007]

许美凤,国雷达,宋盼盼,等. 多卷积神经网络模型融合的皮肤病识别方法. 2019,53(11):125-130. [doi:10.7652/xjtuxb201911018]

胡怀中,张建博,刘汉青,等. 采用卷积神经网络的电站锅炉燃烧效率建模方法. 2019,53(10):10-15. [doi:10.7652/xjtuxb201910002]

张康,陶建峰,覃程锦,等. 随机丢弃和批标准化的深度卷积神经网络柴油机失火故障诊断. 2019,53(8):159-166. [doi:10.7652/xjtuxb201908021]

贾晓芬,郭永存,柴华荣,等. 深立井井壁图像的卷积神经网络去噪方法. 2019,53(6):117-124. [doi:10.7652/xjtuxb201906016]

余星达,陈文杰,王鼎,等. 非接触式身份识别的深度学习算法. 2019,53(4):122-127. [doi:10.7652/xjtuxb201904018]

姜慧明,秦贵和,邹密,等. 改进U型卷积网络的细胞核分割方法. 2019,53(4):100-107,121. [doi:10.7652/xjtuxb201904015]

王晓芳,齐春. 一种运用显著性检测的行为识别方法. 2018,52(2):24-29,44. [doi:10.7652/xjtuxb201802004]

(编辑 陶晴)