DOI:10.19651/j.cnki.emt.2107608

人体动作识别的特征级融合 LSTM-CNN 方法研究*

杨万鹏 李 擎 雷 明
(北京信息科技大学 自动化学院 高动态导航技术北京市重点实验室 北京 100192)

摘 要: 近年来,深度学习方法在人体动作识别有着良好的表现,其利用陀螺仪和加速度计等可穿戴传感器获得的时间序列数据,经过预处理和数据级融合之后进行训练分类。针对数据级融合方法对多传感器的识别有一定局限性的问题,提出了一种特征级融合的 LSTM 和 CNN 方法。该方法将独立的传感器数据依次接入到 LSTM 层和卷积组件层用于特征提取,之后汇聚起多传感器的特征再进行动作分类。该方法在 3 个公开数据集 UCI-HAR、PAMAP2 和 OPPORTUNITY 上分别取得的平均 F1 分数为 96.06%、96.17%和 94.44%。实验结果表明,所提出的方法在多传感器识别人体动作上有较好的精度。

关键词: 人体动作识别;特征级融合;深度学习;多传感器
中图分类号: TP212 **文献标识码:** A **国家标准学科分类代码:** 510.4

Research on feature-level fusion LSTM-CNN method for human activity recognition

Yang Wanpeng Li Qing Lei Ming
(Beijing Key Laboratory of High Dynamic Navigation Technology, School of Automation, Beijing Information Science and Technology University, Beijing 100192, China)

Abstract: In recent years, deep learning methods have performed well in human activity recognition. They use time series data obtained by wearable sensors such as gyroscopes and accelerometers to perform training and classification after preprocessing and data-level fusion. This paper proposes a feature-level fusion method of LSTM and CNN in order to solve the problem that the data-level fusion method has certain limitations in the recognition of multiple sensors. This method connects the independent sensor data to the LSTM layer and the convolutional component layer in turn for feature extraction, and then gathers the features of multiple sensors for action classification. The average F1 scores of this method on the three public data sets UCI-HAR, PAMAP2 and OPPORTUNITY is 96.06%, 96.17% and 94.44% respectively. Experimental results show that the method proposed has better accuracy in multi-sensor recognition of human movements.

Keywords: human activity recognition; feature-level fusion; deep learning; multi-sensor

0 引 言

人体动作识别(human activity recognition, HAR)是通过对人类行为及环境的一系列观察和分析,推断出人体当前的行为和目标。目前,该技术已广泛应用于行人导航[1]、步态分析[2]、手势识别[3]等领域。主要有两种类型的 HAR:基于视频的 HAR 和基于传感器的 HAR[4]。基于视频的 HAR 是通过捕捉图像、视频或监控摄像头的人体动作来识别人类日常动作;而基于传感器的 HAR 则是关注于加速计、陀螺仪、磁强计、蓝牙、声音传感器等智能传感器数据,通过传感器有规律的数据变化来判断人类具体动作。随着可穿戴传感器的引入,传感器可以放置在身体的不同位置收集数据,以推断人类动作类别。

单一传感器来检测人类动作不是很可靠,许多研究提出信息融合策略,结合多个传感器提高鲁棒性,可靠性,降低了数据误分类的概率。传感器融合一般分为数据级融合、特征级融合和决策级融合[5]。其中数据级融合是将原始数据在输入分类器之前进行融合,但是这必然会丢失一些运动的独有特征。决策级则是将分类结果进行融合,但是深度学习方法在识别人体动作分类上已经有很高的精

收稿日期:2021-07-12
* 基金项目:国家自然科学基金(61971048,61771059)项目资助

度，则此方法也见效甚微。而特征级融合的方法既能有效提取传感器的独立特征，也能达到很高的精度。

HAR 的处理流程首先是对传感器进行数据采集，然后再对数据进行预处理、数据分割、提取动作特征，最后对动作进行分类。传统的机器学习方法在推断动作方面发挥了出色的表现。其中包括支持向量机（SVM）[6]、K-近邻（KNN）[7]、随机森林（RF）[8]和隐马尔科夫模型（HMM）[9]等。但是传统的机器学习还有很多局限性，特征提取总是依靠手工方法去精心设计与选择，这严重依赖于人类经验或领域知识。此外，手工设计的特征不能代表复杂活动的显著特征，并且选择最佳特征也是及其耗时[9]。

与传统机器学习方法不同，深度学习在很大程度上减轻了设计特征的工作量，通过训练端到端的神经网络，可以自动提取特征，而不需要去手动设计[10]。并且提取的特征可以表示多模态数据之间的局部联系和数据的时间相关性。因此，深度学习是 HAR 的理想方法，并在现有的工作中得到了广泛的探索：Wan 等[11]提出基于智能手机惯性加速度计的深度学习框架，利用三层卷积池化来提取局部特征实现人类动作分类，此外该作者还使用 LSTM 和 BiLSTM 等方法进行比较；Teng 等[12]探究轻量级 CNN 滤波器是否适用于深部 HAR 任务。将滤波器的类似于乐高积木的方法堆叠一起。Maitre 等[13]将传感器的原始数据和经典特征接入到卷积层、密集层和连接层，实现特征级融合；Mahmud 等[14]使用两层 LSTM 网络从不同的传感器提取时间特征，将这些特征聚集起来送入到由 3 个连续的 LSTM 层组成的全局特征优化器网络，优化聚集的特征向量。除了单一的深度学习方法，现在的 HAR 研究趋向多种深度学习网络结合的方法，而 CNN 与 RNN 的结合是现在普遍而广泛的方法。Chen 等[15]提出了基于 CNN 和 LSTM 的复杂人类动作识别方法，该方法对不同的传感器数据设计了特定的卷积子网络结构，并将所有子网络的输出通过全连接层和卷积层提取融合特征，再使用两层 LSTM 网络来提取潜在动作特征；Xia 等[16]提出了将 LSTM 和 CNN 结合的深度神经网络，该模型的原始数据被输入到两层 LSTM，然后再输入卷积层，使得 LSTM 能够根据学习到的参数学习不同时间尺度上的时间动态，从而获得更好的精度。并且使用全局平均池化层代替卷积后的全连通层在保持高识别率的同时大大降低了模型参数；王震宇等[17]提出了一个 DeepConvGRU 模型，它使用卷积神经网络和门控递归单元作为学习模型，执行自动特征提取和动作分类。

传统的 CNN 和 LSTM 方法虽然能提取动作的显著空间特征，但是时间特征的相关性就大大减弱。并且采用多传感器数据融合的方式，使来自不同传感器的特征不能提取的很清晰。由此本文提出了特征级融合 LSTM-CNN 的网络，该网络由 LSTM 层、卷积组件层和特征级融合层组成，该网络能有效地提取传感器测试的不同动作类型时间序列数据的特征，并且在 3 个可公开访问的数据集（UCI-HAR、PAMAP2 和 OPPORTUNITY）上进行性能测试。

1　人体动作数据集与数据预处理

1.1　数据集介绍

UCI-HAR 数据集[18]：该数据集是由 30 名 19～48 岁的参与者通过佩戴了一款腰间嵌有惯性传感器的智能手机进行采集的，动作类别是根据实验录像人工标注的。实验主要进行 6 项人体动作（躺、站、坐、下楼、上楼和走路）和 6 项动作转换（站到坐、坐到站、坐到躺、躺到坐、站到躺、躺到站）。其中手机的惯性传感器主要是陀螺仪和加速度计。实验人员将加速度计输出分成人体加速度和重力分量两部分。所以原始时间序列数据包括陀螺仪、人体加速度和总加速度。在本文主要分类 6 项人体动作，不考虑动作转换。模型输入为上述 3 类数据，其中每类数据有 3 个维度。所以在此数据集分别对这 3 类数据进行特征提取。

PAMAP2 数据集[19]：该数据集记录了 9 名参与者包括 1 名女性和 8 名男性的 18 项日常活动。其中包括 12 项协议动作（行走、跑步、真空清洁、跳绳等）和 6 项可选动作（看电视、电脑工作、叠衣服等）。参与者的动作数据是从一个心率监视器和佩戴手部、胸部和脚的 3 个惯性传感器捕获的，其中惯性传感器包括陀螺仪、加速度计和磁力计。传感器的采样频率为 100 Hz。该数据集共收集 54 维数据，其中加速度计数据分别有 6 g 和 16 g 两种量程。在本文主要分类 12 项协议动作，选择 3 个惯性器件的陀螺仪、16 g 量程加速度计和磁力计共计 27 维数据作为模型输入。

OPPORTUNITY 数据集[20]：该数据集是由 4 名参与者在一个安装了大量传感器的环境里收集的一系列的日常动作组成。该环境中的传感器已经整合到环境、物体和参与者身体上，这些传感器的采样频率为 30 Hz。本文使用该数据集中的手势识别任务，这是一个 18 类的分类问题。在收集数据时需要一个人需要进行 5 次日常活动（ADL）和 1 次重复固定活动（Drill）。其中参与者佩戴传感器包括身上的 5 个惯性传感器和 12 个三轴蓝牙加速度计，脚上的 2 个惯性立方体。其中惯性传感器共 9 维数据，由加速度计、陀螺仪和磁力计组成。1 个惯性立方体有 16 维数据。所以总计 113 维数据。由于传感器较多，所以在本文主要将此数据集输入分为 8 类，身上的惯性测量单元分为 5 类，脚上的惯性立方体分为两类，身体上的 12 个蓝牙加速度计作为一类数据。

1.2　数据预处理

由于无线传感器的不可靠性，数据集中存在缺失数据，本文通过线性插值来填充缺失的值。此外，在所有数据输入到深度模型之前都要归一化到[0，1]的范围。数据集中的异常数据被设置为边界值或直接切断。由于原始传感器数据是按时间顺序排列的，所以无法直接对整个时间序列进行模型输入。因此采用滑动窗口的方法进行数据分割，

所有窗口有 50%的重叠率。对于 UCI-HAR 数据集窗口时间 2.56 s,窗口长度 128。对于 PAMAP2 数据集窗口时间 1.28 s,窗口长度 128。对于 OPPORTUNITY 数据集窗口时间 0.8 s,窗口长度 24。

2　模型架构

特征级融合 LSTM-CNN 模型架构如图 1 所示。单一传感器的数据经过预处理后被输入到两层的 LSTM 网络中,它用于提取序列的时间特征。接下来是 3 个卷积组件,每个卷积组件由卷积层(convolutional)、最大池化层(max pooling)、批量标准化层(batch normalization, BN)和 Dropout 层组成,其作用是提取序列的空间特征。之后再接入一个全局平均池化层(global average pooling, GAP),以上作为一个单一传感器的特征提取。多个传感器特征提取完毕后被输入到特征级融合层进行特征融合,特征级融合层之后连接批处理批量标准化层,紧接着是使用 Softmax 激活功能并对输入特征进行分类的全连接层。

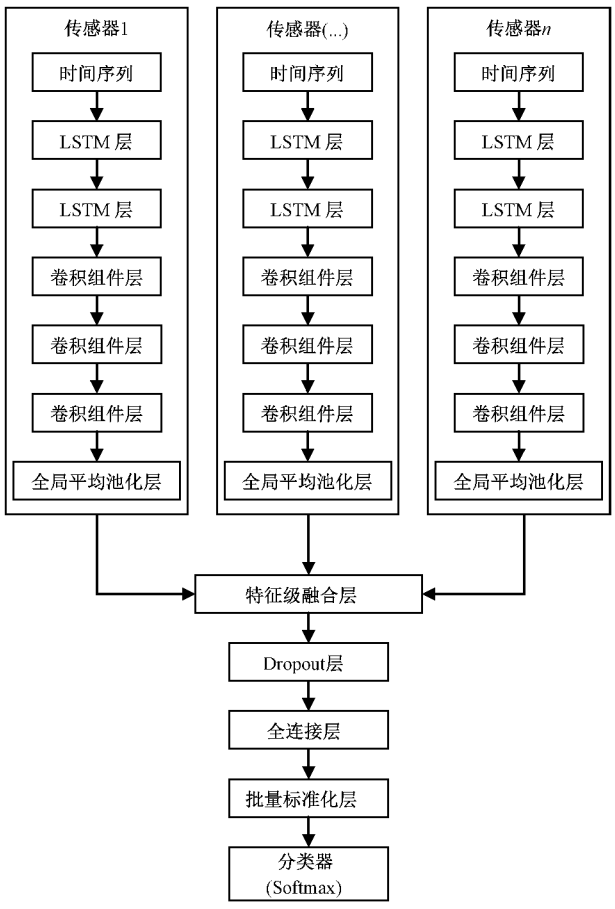

图 1　特征级 LSTM-CNN 模型架构

2.1　LSTM 层

循环神经网络(RNN)可以提取时间序列中时间特征,但是 RNN 在每个时间步长会丢失一些信息,长时间之后几乎记忆不了最初的输入部分[21]。所以一般使用长短时记忆网络(LSTM)记忆细胞代替循环单元扩展了 RNN,主要优点是不仅可以存储和输出信息,还可以在长时间尺度上简化了对时间关系的学习。

LSTM 主要的作用是针对网络隐藏层对其前向及后向传播进行网络计算,以隐藏层神经元来模拟传感器数据的时间依赖性[22]。基于 LSTM 的深度递归神经网络(DRNN)模型可以利用自动分层从时间序列数据提取深度学习特征。在 LSTM 一个神经单元中状态被分为短期状态和长期状态 c_t,两种状态连接 4 个不同的全连接层,主要层是输出 g_t 的层和 3 个门控制器,包括输入门 i_t,遗忘门 f_t 和输出门 o_t,以控制记忆细胞的行为。每个 LSTM 单元的激活情况由以下公式计算。

$$i_t = \sigma(\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i) \tag{1}$$

$$f_t = \sigma(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f) \tag{2}$$

$$o_t = \sigma(\boldsymbol{W}_{xo}\boldsymbol{x}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o) \tag{3}$$

$$g_t = \sigma_h(\boldsymbol{W}_{xg}\boldsymbol{x}_t + \boldsymbol{W}_{hg}\boldsymbol{h}_{t-1} + \boldsymbol{b}_g) \tag{4}$$

$$c_t = f_t c_{t-1} + i_t g_t \tag{5}$$

其中,$\boldsymbol{W}_x$ 是 4 层与输入向量 $\boldsymbol{x}_t$ 的连接权重矩阵;$\boldsymbol{W}_h$ 是与短期状态 $\boldsymbol{h}_{t-1}$ 的连接权重;$\boldsymbol{b}$ 是 4 层的偏置项矩阵;σ 表示 sigmoid 激活函数;σ_h 表示 tanh 激活函数。本文架构中,传感器预处理之后的数据首先被连接到有 64 个神经元的两层 LSTM 中。

2.2　卷积组件层

时序序列经过 LSTM 层处理之后,再接入到卷积组件层。其中 1 个卷积组件包括卷积层、批量标准化、最大池化层和 Dropout 层,其结构如图 2 所示。在本文模型连续接入 3 组卷积组件层,其中 3 个卷积层的卷积核数量分别为 64、128 和 256,卷积核尺寸分别为 3、6 和 9,卷积滑动步长均为 1。3 个池化滤波器的尺寸均为 2×1,滑动步长为 1。在进行特征融合之前要对卷积输出的特征向量进行平铺处理。即在 3 层卷积组件层之后接入全局平均池化层。

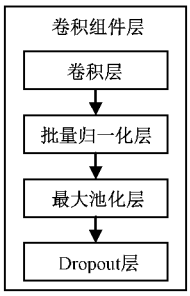

图 2　卷积组件层

CNN 通过一系列卷积运算提取出人类的动作信号特征。它既可以获得动作信号的局部基本特征,又可以在复杂动作中提取不同的动作模式特征[23]。CNN 为输入信号的不同分类分配不同的特征学习策略,可有效实现对传感器的时间序列信号中的特征提取。

卷积层是 CNN 中最重要的组成部分,利用卷积核对

输入进行卷积。对于传感器信号这种一维时间序列，一般采用一维卷积核对时序数据进行卷积操作[24]。其中卷积核可以看作一个过滤器，可以剔除异常值、过滤数据和检测特别行为。每个卷积层的输出是一组特征映射，通过对卷积结果加上偏差然后被 ReLU 激活函数激活，如下所示：

$$f(x)=\max(0,x) \tag{6}$$

批量标准化层(BN)[25]作用是调整各层激活值分布使其拥有适当的广度。在训练过程中，由于上层权重参数的不断更新，每一层输入数据的分布都会不断变化。因此，需要改变权值参数来适应这种新的分布，这将导致网络训练困难和收敛速度减慢。为了解决这个问题，在间隙层之后添加了一个 BN 层来加速模型的收敛性。BN 层对训练样本的输入数据进行归一化和重构，既保证了前一层输出的稳定性，又提高了训练的速度和精度。紧跟 BN 层之后的是最大池化层。最大池化层是为了减小了特征映射，通过计算给定区域的最大值保留主要特征，从而简化计算，避免过拟合。

在传统 CNN 模型中都会在卷积之后加入完全连接层，但是完全连接层容易过度拟合并且权值参数可能占比较大，从而阻碍了整个网络的泛化能力。所以本文使用一个全局平均池化层(GAP)[26]来替代完全连通层。具体做法是对卷积后的特征向量进行平均池化得到结果向量，其优点是全局平均池化不需要优化额外的模型参数，因此模型大小和计算量较全连接大大减少，并且可以避免过拟合。

2.3　特征融合与输出层

多个传感器分别经过以上架构提取特征后，就被送入到特征融合层。特征融合层使用的是密集卷积网络(DenseNet)中介绍的 concatenate 操作[27]，将特征提取后得到的平铺向量进行串联聚合，其结构如图 3 所示。特征融合层可以实现对数据量的压缩和不同类型传感器的融合，这样可以使网络更紧密更有效。在特征融合后将融合向量送入到输出层。

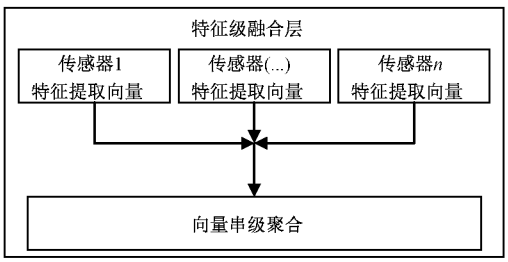

图 3　特征级融合层

输出层是由全连接层、批量标准化层和 Softmax 分类器组成。全连接层的每个节点都连接到上层的节点，以便可以合并从上层提取的特征，弥补了 GAP 层在这方面的不足。最后的分类器使用的是 Softmax 分类器，它将输出的上一层转换为一个概率向量，其值代表当前样本所属的类的概率。表达式公式如下：

$$\boldsymbol{P}_j=\frac{\mathrm{e}^{a^j}}{\sum_{k=1}^{N}\mathrm{e}^{a^k}} \tag{7}$$

其中，e 为指数函数；N 为类数同时也是全连通层的输出向量个数；$\boldsymbol{a}$ 为全连通层的输出向量，a^j 是输出向量的值。

2.4　模型的训练

在本文中，算法是使用带有 TensorFlow 后端的 Keras 框架实现的。模型采用的是交叉熵损失函数，通过计算预测值和真实值的误差来评估概率分布之间的差异。使用 Adam 优化器对模型进行训练，该优化器具有高效计算、所需内存少等优点。学习率采用回调函数的方式自适应设置，初始学习率设置为 0.001 当检测到学习率停滞下降学习率达到最好的效果。模型超参数如表 1 所示。

表 1　模型超参数

超参数	选取的值
循环层神经元	64
卷积核数量	64、128、256
卷积核尺寸	3、6、9
池化尺寸	2
Dropout 参数	0.3
优化器	Adam
迭代次数	100
批处理大小	160
学习率	0.001(最大)、0.000 01(最小)
输入向量维度	128(UCI-HAR) 128(PAMAP2) 24(OPPORTUNITY)
传感器通道	3(UCI-HAR) 3(PAMAP2) 9(OPPORTUNITY)

3　实验结果及分析

本文所有的实验都是在 Windows 电脑上运行的，电脑的 CPU 为 Intel(R) Core(TM) i7-10750H，内存为 64 GB，GPU 为 NVIDIA GeForce RTX2060 显卡，8 G 显存。实验将 3 个数据集样本的 70% 划分为训练集，30% 划分为测试集。

3.1　模型评估

常用的分类模型评估指标包含准确率(Accuracy)、精度(Precision)、召回率(Recall)、F1 分数(F1 score)和混淆矩阵。但在自然环境下，收集人类动作数据时经常面临数据不平衡的情况[28]。上述提到的 PAMAP2 和 OPPORTUNITY 都是不平衡的数据集，整体的准确率难以作为评估模型性能的精确度量。F1 分数同时考虑了假阴性和假阳性，结合

精度和召回率,通过样本比例对类进行加权可消除类中的不平衡。本文主要采用准确率和 F1 分数来评价特征级 LSTM-CNN 模型,具体公式如下:

$$Accuracy = TP + TN/TP + TN + FP + FN \quad (8)$$

$$Precision = TP/TP + FP \quad (9)$$

$$Recall = TP/TP + FN \quad (10)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

式中:TP、FP 分别表示真阳性和假阳性的数量;TN、FN 为真阳性和假阴性的数量。

由于深度学习模型的训练受到优化算法导致的随机现象的影响,所以在每个数据集上分别进行 30 次训练,绘制 3 个数据集 F1 分数。在 UCI-HAR、PAMAP2 和 OPPORTUNITY 3 个数据集上平均准确率和 F1 分数及其方差如表 2 所示。由数据集方差可见当运动分类类别增多,对数据的分类结果波动较大。由数据集上的平均迭代一次的训练时间可知,当训练集样本和分类动作数量增多时,所消耗的训练时间也在相应的增多。将 F1 分数 30 次的结果进行绘制如图 4 所示。在上述 3 个数据集上最大 F1 分数能达到 96.84%、97.53%和 96.34%。在 3 个数据集上即使最低的 F1 分数也能达到 95.44%、92.85%和 91.24%,表现效果依然很好。

表 2　平均准确率 F1 分数和训练时间

性能指标	UCI-HAR	PAMAP2	OPPORTUNITY
平均准确率/%	96.11	96.28	94.47
准确率方差	0.08	0.96	0.98
平均 F1 分数/%	96.06	96.17	94.44
F1 分数方差	0.08	0.95	0.98
训练集样本	7 352	18 820	315 435
分类动作	6	12	18
平均训练时间/s	4.083	5.069	7.049

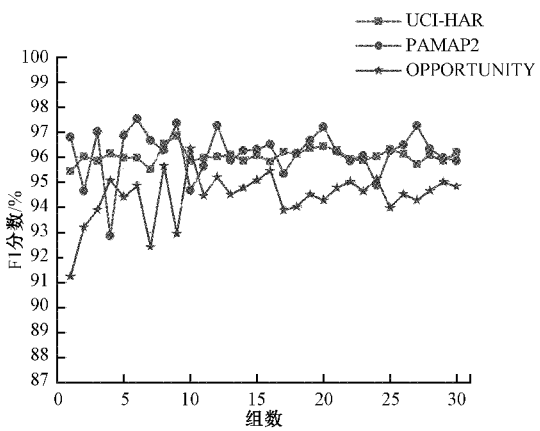

图 4　F1 分数曲线

混淆矩阵是评价分类效果的可视化方法,它提供了模型总体分类率。本文在方差范围内,选取与平均值最接近的一次实验结果绘制 3 个数据集的混淆矩阵。

从 UCI-HAR 数据集混淆矩阵可知,如图 5 所示,该模型在此数据集上识别良好。躺和下楼都能识别到 100%。而其中走与上楼和坐与站的活动类似,所以两者易混淆程度较大。

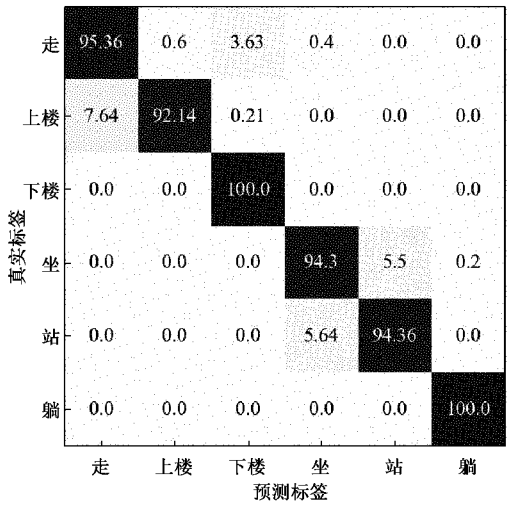

图 5　UCI-HAR 数据集混淆矩阵

从 PAMAP2 数据集混淆矩阵可知,如图 6 所示,多数动作均能识别到 90%以上,其中站立和越野走准确率均能达到 100%。但是跑步的识别效果较差,其中大多数被识别成骑车。在复杂运动中,跑步属于其他动作的从属动作,所以经常会被识别错误。

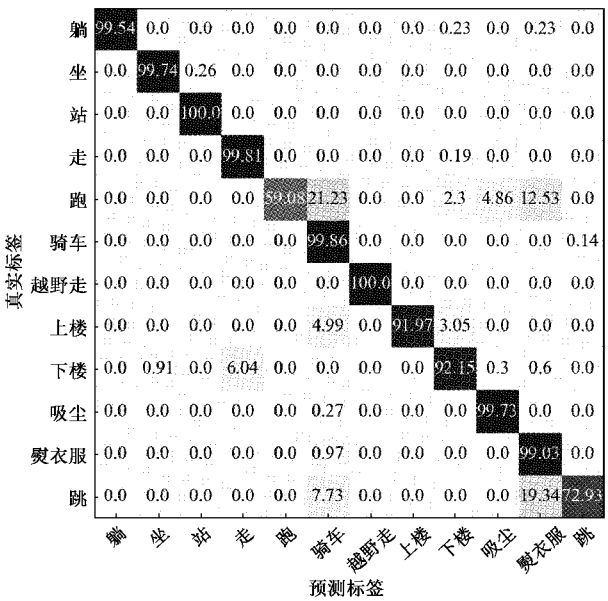

图 6　PAMAP2 数据集混淆矩阵

从 OPPORTUNITY 数据集混淆矩阵可知,如图 7 所示,数据识别准确率多数集中在 80%～92%之间。其中大多数的错误来自于无动作。这是因为在长时间记录中除了

有意义的动作都被归结于无动作。剩下的错误率集中在相对相似的手势之间，例如“打开洗碗机”，“关闭洗碗机”或者“打开抽屉 1”、“打开抽屉 2”。这是因为这些手势由相同类型传感器识别，只是识别顺序不同。

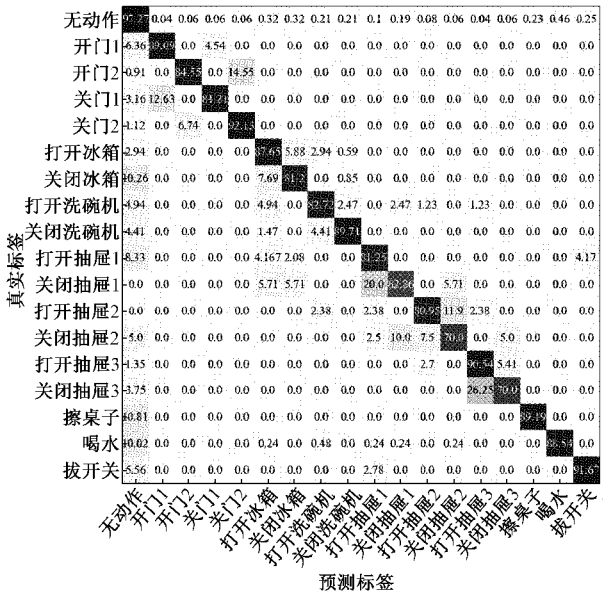

图 7　OPPORTUNITY 数据集混淆矩阵

3.2　不同传感器的性能对比

人体动作识别使用了多种佩戴在身体各处的传感器，为了分析使用各种传感器提取的时间序列数据特征对分类结果的影响，本文展示了特征级 LSTM-CNN 在识别 OPPORTUNITY 数据集上使用不同传感器对分类动作的性能对比结果。OPPORTUNITY 数据集上的传感器可分为 3 类，包括佩戴参与者身体惯性测量单元、脚部的惯性立方体、集成到环境中的无线蓝牙加速度计。如表 3 所示，脚步佩戴的传感器和集成到环境中的传感器产生的数据分类结果的平均 F1 分数只能达到 71.57%和 81.76%，说明识别动作的准确率并不高。相对而言，仅身体佩戴和脚部与身体同时佩戴的传感器均能达到更高的分类效果，分别可达到 90.52%和 92.22%，而佩戴全套设备的传感器可达到 94.44%。这说明传感器的数量和集成到身体的位置与分类效果有很大关系，因此相比于单一传感器，结合多传感器的数据，从多角度对动作进行划分能够产生更好的分类效果。

表 3　不同传感器的性能对比

性能指标	脚部佩戴	身体佩戴	集成环境中	佩戴脚部和身体	全套设备
传感器数量	2	5	12	7	19
输入维度	32	45	36	77	113
F1 分数/%	71.57	90.52	81.76	92.22	94.44

3.3　网络结构对比

为了验证特征级融合的 LSTM 和 CNN 模型的有效性及优越性，本文将特征级 LSTM-CNN 模型与下述两种模型进行对比。一种是采用数据级融合的方式，其训练模型与本文提出的相同，该模型称为数据级 LSTM-CNN。另外一种仍是采用特征级融合的方式，但是模型结构是先接入卷积组件层在接入 LSTM 层。因为 LSTM 层有平铺效果，所以去掉了 GAP 层，其中模型超参数不发生改变，该模型称为特征级 CNN-LSTM。所有结果均以准确率进行验证，以确保以下比较结果的公平性和一致性。

上述深度模型在 3 个公开数据集的评估结果，如图 8 所示。在 3 个数据集中特征级 LSTM-CNN 比数据级 LSTM-CNN 分别高于 4.45%、15.55%和 2.08%。由此可见多传感器分别进行特征提取的方法是优于数据融合的方法。传感器相对独立的特征提取，能够有效地保留其独有特征。虽然特征级 LSTM-CNN 只是略高于特征级 CNN-LSTM。可以说明先采用 LSTM 对输入进行提取特征，能有效提取传感器之间的联系。

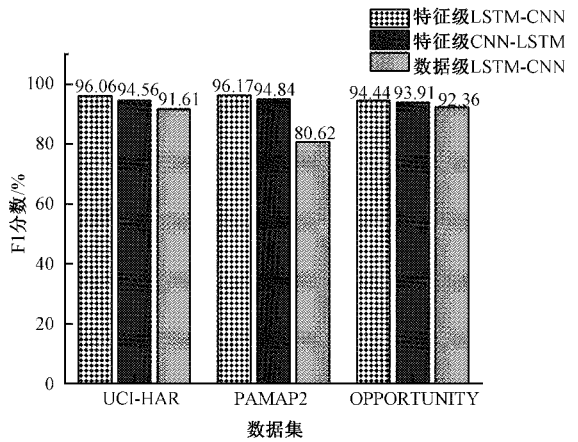

图 8　网络结构对比

3.4　与其他模型对比

为了验证本文提出模型的优越性，与使用过 UCI-HAR、PAMAP2 和 OPPORTUNITY 3 个数据集的论文中给出的模型进行了比较。其中模型有上述提到 Wan 等提出的 CNN 和 LSTM 模型；Francisco 等[29]提出的 DeepConvLSTM 模型；王震宇等提出的 DeepConvGRU 模型；Xia 等提出的 LSTM-CNN 模型。模型根据准确率和 F1 分数进行比较。如表 4 所示，给出了比较结果。从表 4 中可以看出，单一的深度学习方法例如 CNN 和 LSTM 的评价指标都要低于此两种方法的结合。而采用 CNN 和 RNN 结合的两种方法中，在 OPPORTUNITY 数据集上 DeepConvLSTM 的 F1 分数比本文的方法要低 2.94%；DeepConvGRU 的 F1 分数要低于本文方法 3.14%。而采取数据融合的 LSTM 和 CNN 结合的架构，对于其他模型性能均有提升，本文提出的特征级融合方法要较好于该模型。

表 4 与其他模型对比结果 %

模型	UCI-HAR		PAMAP2		OPPORTUNITY	
	准确率	F1 分数	准确率	F1 分数	准确率	F1 分数
CNN	92.70	92.90	91.16	91.00	—	—
LSTM	88.99	89.01	85.34	85.86	—	—
DeepConvLSTM	—	—	—	—		91.50
DeepConvGRU	93.80	—	—	—	—	91.30
LSTM-CNN	95.78	—	—	—	—	92.63
特征级融合 LSTM-CNN	96.11	96.06	96.28	96.17	94.47	94.44

4 结 论

本文提出的特征级融合的 LSTM 和 CNN 相结合的深度学习模型,该网络的主要特点主要有两点。首先是网络的架构,本文采用 LSTM 加 CNN 的模式来搭建模型,其中两层 LSTM 和三层卷积组件的组合能有效地提取数据的深度特征。其次本文采用特征融合的方式来对运动状态进行分类,这样可以使来自多传感器的数据能保留其独有的特征。

本文通过在 3 个公开数据集对该模型进行验证,在每个数据集上都取得了良好的效果。本文模型识别精度高于数据级融合的 LSTM-CNN 和特征级融合的 CNN-LSTM。相对比数据级融合,特征级融合能有效提取单一传感器的关系。而对于 CNN-LSTM,LSTM-CNN 能够有效提取传感器的时间相互关系。而与其他论文中的模型对比也说明本文提出的模型有良好的泛化性能。从结果中还观察到,该模型在包含各种动作的数据集上表现良好。

参考文献

[1] LI C, SU Z, LI Q, et al. An indoor positioning error correction method of pedestrian multi-motions recognized by hybrid-orders fraction domain transformation[J]. IEEE Access, 2019, 7: 11360-11377.

[2] HAMMERLA N Y, HALLORAN S, PLOETZ T. Deep convolutional and recurrent models for human activity recognition using wearables[J]. Journal of Scientific Computing, 2016, 61(2): 454-476.

[3] KIM Y, TOOMAJIAN B. Hand gesture recognition using micro-doppler signatures with convolutional neural network[J]. IEEE Access, 2016, 4: 7125-7130.

[4] WANG J, CHEN Y, HAO S, et al. Deep learning for sensor-based activity recognition: A survey[J]. Pattern Recognition Letters, 2019, 119: 3-11.

[5] NOORI F M, RIEGLER M, UDDIN M Z, et al. Human activity recognition from multiple sensors data using multi-fusion representations and CNNs[J]. ACM Transactions on Multimedia Computing Communications and Applications, 2020, 16(2): 1-19.

[6] GARCIA-GONZALEZ D, RIVERO D, FERNANDEZ-BLANCO E, et al. A public domain dataset for real-life human activity recognition using smartphone sensors[J]. Sensors(Basel, Switzerland), 2020, 20(8): 2200.

[7] PATRO S, MISHRA B K, PANDA S K, et al. A hybrid action-related k-nearest neighbour(HAR-KNN) approach for recommendation systems[J]. IEEE Access, 2020, 8: 90978-90991.

[8] PATHAN N S, TALUKDAR M, QUAMRUZZAMAN M, et al. A machine learning based human activity recognition during physical exercise using wavelet packet transform of PPG and inertial sensors data[C]. 2019 4th International Conference on Electrical Information and Communication Technology(EICT), 2019: 1-5.

[9] SIDDIQI M H, ALRUWAILI M, ALI A, et al. Human activity recognition using gaussian mixture hidden conditional random fields[J]. Computational Intelligence and Neuroscience, 2019:8590560-8590560.

[10] NWEKE H F, TEH Y W, AL-GARADI M A, et al. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges[J]. Expert Systems with Application, 2018, 105: 233-261.

[11] WAN S, QI L, XU X, et al. Deep learning models for real-time human activity recognition with smartphones[J]. Mobile Networks & Applications, 2020, 25(2): 743-755.

[12] TENG Q, WANG K, ZHANG L, et al. The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition[J]. IEEE Sensors Journal, 2020, 20(13):7265-7274.

[13] MAITRE J, BOUCHARD K, GABOURY S. Alternative deep learning architectures for feature-level

fusion in human activity recognition [J]. Mobile Networks and Applications, 2021, DOI: 10.1007/s11036-021-01741-5.

[14] MAHMUD T, AKASH S S, FATTAH S A, et al. Human activity recognition from multi-modal wearable sensor data using deep multi-stage LSTM architecture based on temporal feature aggregation[C]. 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems(MWSCAS), IEEE, 2020: 249-252.

[15] CHEN L, LIU X, PENG L, et al. Deep learning based multimodal complex human activity recognition using wearable devices [J]. Applied Intelligence, 2020,51(6): 4029-4042.

[16] XIA K, HUANG J, WANG H. LSTM-CNN architecture for human activity recognition[J]. IEEE Access, 2020, 8:56855-56866.

[17] 王震宇，张雷. 基于深度卷积和门控循环神经网络的传感器运动识别[J]. 电子测量与仪器学报，2020，34(1):1-9.

[18] ANGUITA D, GHIO A, ONETO L, et al. A public domain dataset for human activity recognition using smartphones [C]. Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2013: 437-442.

[19] REISS A, STRICKER D. Introducing a new benchmarked dataset for activity monitoring [C]. International Symposium on Wearable Computers, IEEE, 2012: 108-109.

[20] CHAVARRIAGA R, SAGHA H, CALATRONI A, et al. The opportunity challenge: A benchmark database for on-body sensor-based activity recognition[J]. Pattern Recognition Letters, 2013, 34(15): 2033-2042.

[21] INCE I F. Performance boosting of scale and rotation invariant human activity recognition (HAR) with LSTM networks using low dimensional 3D posture data in egocentric coordinates[J]. Applied Sciences, 2020, 10(23): 8474.

[22] LI X, WANG Y, ZHANG B, et al. PSDRNN: An efficient and effective HAR scheme based on feature extraction and deep learning[J]. IEEE Transactions on Industrial Informatics, 2020, 16(10): 6703-6713.

[23] ZHANG H, XIAO Z, WANG J, et al. A novel IoT-perceptive human activity recognition(HAR) approach using multihead convolutional attention [J]. IEEE Internet of Things Journal, 2020, 7(2): 1072-1080.

[24] ZHOU B, YANG J, LI Q. Smartphone-based activity recognition for indoor localization using a convolutional neural network[J]. Sensors, 2019, 19(3): 621.

[25] MÜNZNER S, SCHMIDT P, REISS A, et al. CNN-based sensor fusion techniques for multimodal human activity recognition[C]. Acm International Symposium. ACM, 2017:158-165.

[26] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]. CVPR, IEEE Computer Society, 2016: 2921-2929.

[27] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks[J]. IEEE Computer Society, 2017: 2261-2269.

[28] RONAO C A, CHO S B. Human activity recognition with smartphone sensors using deep learning neural networks [J]. Expert Systems with Applications, 2016, 59:235-244.

[29] FRANCISCO O, DANIEL R. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition [J]. Sensors, 2016, 16(1):115.

作者简介

杨万鹏，硕士研究生，主要研究方向为智能导航、深度学习。

E-mail:18210540798@163.com

李擎(通信作者)，博士，教授，主要研究方向为智能导航、自主系统。

E-mail:liqing@bistu.edu.cn

雷明，硕士研究生，主要研究方向为惯性导航、智能导航。

E-mail:928080912@qq.com