



计算机工程
Computer Engineering
ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目: 基于卷积长短时记忆网络的人体行为识别研究
作者: 孙彦玺, 赵婉婉, 武东辉, 陈继斌, 仇森
DOI: 10.19678/j.issn.1000-3428.0060938
网络首发日期: 2021-05-13
引用格式: 孙彦玺, 赵婉婉, 武东辉, 陈继斌, 仇森. 基于卷积长短时记忆网络的人体行为识别研究. 计算机工程. <https://doi.org/10.19678/j.issn.1000-3428.0060938>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。



基于卷积长短时记忆网络的人体行为识别研究

孙彦玺¹, 赵婉婉¹, 武东辉^{1*}, 陈继斌¹, 仇 森²

(1. 郑州轻工业大学, 建筑环境工程学院, 郑州 450002; 2. 大连理工大学, 控制科学与工程学院, 辽宁 大连 116024)

摘 要: 人体行为识别 (HAR) 是模式识别、普适计算和人机交互等领域的重要研究课题, 具有广泛的社会应用价值, 利用深度学习网络自动提取数据的深层特征和提高对人体行为的分类准确度是人体行为识别的研究热点。考虑到传统机器学习算法严重依赖手工特征提取、模型泛化能力差等问题, 提出了基于时空特征融合的深度神经网络 (CLT-net) 的人体行为识别方法。模型采用卷积神经网络 (CNN) 自动提取人体行为数据的深层次隐含特征, 利用长短时记忆网络 (LSTM) 构建时间序列模型, 学习人体行为特征在时间序列上的长期依赖关系, 最终 softmax 分类器实现对不同人体行为的分类。基于公开 DaLiAc 数据集的实验结果表明, 基于 CLT-net 的人体行为识别方法对 13 种人体行为的总体识别率达到了 97.6%, 优于单一的 CNN 和 LSTM 模型以及传统的 BP 神经网络算法, 展现所提方法用于人体行为识别的有效性。

关键词: 人体行为识别; 深度学习; 卷积神经网络; 长短时记忆网络; 模式识别; 可穿戴传感器

开放科学 (资源服务) 标识码 (OSID):



Research of CNN-LSTM-Based Human Activity Recognition

Sun Yan-xi¹, Zhao Wan-wan¹, Wu Dong-hui^{1*}, Chen Ji-bin¹, Qiu Sen²

(1. College of Building Environment Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China;

2. School of Control Science and Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China)

【Abstract】 Human activity recognition is an important research topic in the fields of pattern recognition, pervasive computing and human-computer interaction. It has a wide range of social application value. Using deep learning network to automatically extract deep features of data and improve the classification accuracy of human activity is a hot research topic of human activity recognition. Considering that traditional machine learning algorithms rely heavily on manual feature extraction and poor generalization ability of the model, a deep learning model based on space-time feature fusion (CLT-net) is proposed for human activity recognition. Convolution neural network (CNN) is used to extract the deep hidden features of human activity data. Long short term memory network (LSTM) is used to construct time series model to learn the long-term dependence of human activity features on the time series. Finally softmax classifier is used to classify different human activities. The experimental results based on public DaLiAc dataset show that the overall recognition accuracy of CNN-LSTM-based human activity recognition for 13 kinds of human activities reaches 97.6%, which is better than the single CNN and LSTM models and the traditional BP neural network algorithm, showing the effectiveness of the proposed method for human activity recognition.

【Key words】 human activity recognition; deep learning; convolutional neural network; long short term memory network; pattern recognition; wearable sensors

DOI:10.19678/j.issn.1000-3428.0060938

0 概述

人体的各种动作、行为及日常活动是构成人们社会生产、生活和交流的重要组成部分, 人体行为识别在人们的日常生活中起着重要的作用, 可以被

广泛应用在医疗康复、智能看护、运动监测和人机交互等众多领域^[1, 2]。根据数据来源的不同, 人体行为识别方法可以被分为两类: 基于视频图像的人体行为识别和基于可穿戴传感器的人体行为识别

基金项目: 国家自然科学基金青年项目 (61803072); 河南省科技攻关项目 (182102210622); 河南省高等学校重点科研项目资助计划 (19A413013); 郑州轻工业大学青年骨干项目 (13501050009)

作者简介: 孙彦玺 (1996—), 男, 硕士研究生, 主研方向为深度学习、人体行为识别; 赵婉婉, 本科、助理工程师; 武东辉, 讲师、博士; 陈继斌, 教授、硕士; 仇森, 讲师、博士。 **E-mail:** 1925194597@qq.com

[3]。基于视频图像的人体行为识别利用图像或视频处理等相关技术,通过对摄像设备获取到的人体运动的图像集合或视频片段进行分析,从而实现对人体行为的识别和特定目标侦测^[4、5]。基于可穿戴传感器的人体行为识别能够从低级别的原始传感器数据中寻求关于人体行为中所蕴含的深层次知识,传感器数据主要由加速度计、陀螺仪和磁力计等传感器采集得到^[6]。随着微电子机械系统(MEMS)和无线通信技术的飞速发展,这些传感器可以被集成在可穿戴设备、智能手机或智能手表中,极大地方便了人们的日常携带,且在实际使用时不会受到场所和周围环境的限制,亦不会给用户带来侵犯个人隐私的威胁^[7],这促进了基于可穿戴传感器的人体行为识别在人们日常生活中的应用。

人体行为识别属于典型的模式识别问题,传统的模式识别问题主要采用了人工神经网络、支持向量机、决策树、朴素贝叶斯、K 最近邻和隐马尔科夫模型等机器学习算法^[8、9]。过去的十几年里,这些机器学习算法在人体行为识别的问题上取得了巨大的进步,但是也存在一些不可回避缺点。例如,利用传统的机器学习算法进行人体行为识别时,需要提前手工提取人体行为数据特征,而手工特征提取则受到特定领域知识和人们已有知识和经验的限制^[10]。其次,提取的一些浅层次特征(如均值、方差、频率等统计信息)只能用于识别人体行为的低级活动(如,站立、行走、跑步等),而很难识别出一些更为复杂和高级的人体行为(如洗盘子、使用吸尘器等),尤其是在当前多模态和高维传感器数据涌现的情况下,这些特征无法有效的处理复杂的活

动和实现对人体活动的准确分类^[11]。

为了进一步提高对基于可穿戴传感器的人体行为识别的分类效果,同时又不依赖于手工特征的提取,本文提出了基于空时特征融合的深度学习方法(CLT-net)的人体行为识别方法,模型充分结合了CNN可以自动提取数据特征和LSTM能够学习时序数据之间的相关性的优点,并在公开数据集上与单一的CNN和LSTM模型及BP神经网络算法进行比较。

1 相关工作

近年来,深度学习技术得到蓬勃发展,且在图像识别^[12]、目标检测^[13]、视频动作识别^[14、15]、自然语言处理^[16]和时间序列预测^[17]等众多领域取得了令人满意的效果。与传统的机器学习算法不同,深度学习网络在很大程度上减轻了研究人员手工提取特征的工作量,模型通过更深层次的网络训练,可以自动提取到更高级别、更有意义的特征,使其在面对复杂的人体行为识别时更显优势^[18]。现阶段最常用于人体行为识别的两种深度学习网络是卷积神经网络(CNN)和长短期记忆(LSTM)。如文献[19]提出了一种基于一维(1D)卷积神经网络(CNN)的方法来进行人体活动识别,该方法对基于智能手机加速度传感器的走路、跑步和静止三种人体活动的识别准确率达到92.71%。文献[20]提出了一种基于2D和3D的卷积神经网络模型学习人体行为数据特征,并结合手工提取的特征用于人体行为识别,在UCI数据集上的实验表明,该方法达到了96.95%的准确度。文献[21]设计了一种长短时记忆递归神经网络(LSTM-RNN)来进行人体活动识别,该方法

可以实现特征的自动提取和融合,分类效果较好。文献[22]提出了一种基于双向长短期记忆(BiLSTM)循环神经网络模型,并采用PCA来降低数据集的维度,在mHealth数据集上的实验结果表明,PCA-BiLSTM模型的准确率达到了97.64%。

本文的其余部分组织如下:在第2节中,简要介绍了我们建立的卷积神经网络(CNN)、长短时记忆网络(LSTM)和基于空时特征融合的深度学习模型(CLT-net)在人体行为时间序列信号上的应用,并给出相应的网络模型。第3节,介绍了我们实验所采用的公开数据集,并详细阐述了实验运行环境以及实验中关键参数的设置。第4节,根据实验结果进行分析。第5节,总结并对未来的工作进行展望。

2 人体行为识别模型

2.1 卷积神经网络模型

虽然深度学习网络在图像分类、人脸识别和自然语言处理等领域已经取得了不错的成绩,但是目前还没有在序列信号分类应用上得到一种公认的结构。因此,本文在参考LeNet-5基础上设计了卷积神经网络(Convolutional Neural Networks, CNN)模型用于人体行为识别,其中与LeNet-5不同的是,本文的CNN模型输入层数据格式为 24×410 的序列,并且在每个最大池化层之后加入批归一化层(Batch Normalization)、激活层(Leaky Relu)。构建的卷积神经网络模型如图1所示,主要包括序列输入层、折叠层、CNN特征提取层、解折叠层、flatten层、全连接层、softmax分类层。

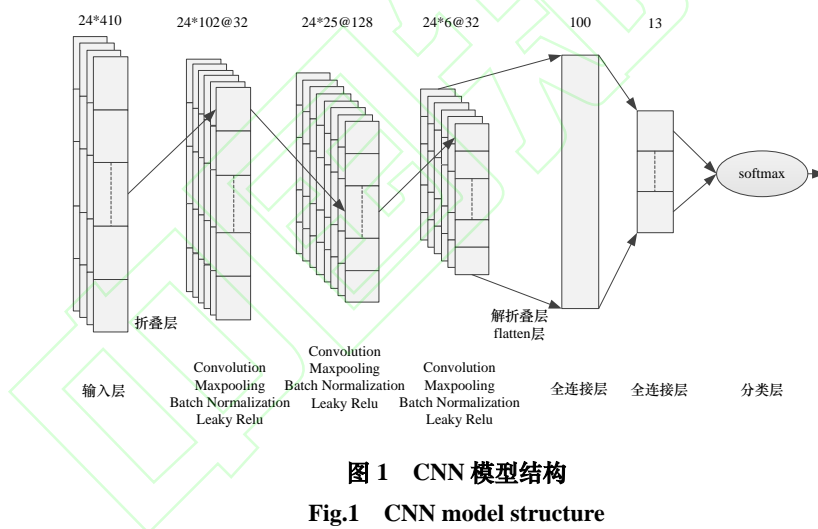


图1 CNN模型结构

Fig.1 CNN model structure

其中,批归一化技术是由Google于2015年提出,它应用于深度神经网络训练不仅可以加快模型的收敛速度,而且可以一定程度缓解了深层网络中“梯度弥散”的问题,从而使得训练的深度学习模型更加稳定。Leaky Relu激活函数是为了解决Relu的输入值为负的时候,输出始终为0,同时一阶导数也始终为0,从而导致神经元参数不更新,也就

是神经元不学习的情况下出现的,其在声学模型(2013)中被首次提出。定义如下:

$$f(x) = \begin{cases} x, & x \geq 0 \\ s * x, & x < 0 \end{cases} \quad (1)$$

式中, s 为一个不小于1的非负数,当取0时,Leaky Relu激活函数退化为Relu函数。

如图1所示,CNN模型共包含3个CNN特征提取层。每个特征提取层的卷积层(Convolution)

可以提取人体行为数据的特征,是 CNN 模型的关键一层;最大池化层 (Maxpooling) 起到了压缩数据、降低维度的作用;批归一化层 (Batch Normalization) 可以对提取出的特征进行归一化处理;非线性激活层 (Leaky Relu) 可以促进批归一化之后特征的映射。全连接层可以减少提取特征的信息损失。softmax 分类层最终实现对人体行为的分类。其中, softmax 分类层的计算公式为:

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}}, i = 1, 2, 3, \dots, K \quad (2)$$

式中, 为提取得到的人体行为数据的特征序列, 为人体行为类别数。softmax 函数的分类结果代表输入样本被划分为每个类别时的所属概率, 且所属概率和为 1。

鉴于卷积神经网络可以自动提取人体行为数据的深层特征, 能够避免手工特征提取带来的诸多缺陷, 因此本文将 CNN 特征提取层作为所提 CLT-net 网络模型的特征提取单元引入。CNN 特征提取层实现特征提取的具体过程如下: 3 个卷积层的卷积核的个数依次设为 32、128、32, 卷积核大小分别设为 (1, 11)、(1, 9)、(1, 7), 步长为 (1, 2), 即

垂直方向步长为 1、水平方向步长为 2, 同时模型采用 same 方式进行 “padding”; 3 个最大池化层的池化核大小均为 (1, 3), 步长为 (1, 2)。当单个样本的输入数据规模为 $24 \times 410 \times 1$ 时, 经过 3 个 CNN 特征提取层得到的人体行为特征序列规模分别为 $24 \times 102 @ 32$, $24 \times 25 @ 128$, $24 \times 6 @ 32$ 。

2.2 长短时记忆网络模型

长短时记忆网络 (Long Short Term Memory Network, LSTM) 是循环神经网络 (RNN) 的改进, 其在 1997 年被 Hochreiter 和 Schmidhuber 提出^[23]。LSTM 网络的核心部分是序列输入层和 LSTM 层, 序列输入层可以将序列或时间序列数据输入网络, LSTM 层可以学习序列数据时间步长之间的长期依赖关系, 很好的解决了 RNN 的梯度消失问题^[24]。鉴于 LSTM 是一种强大的时序信号处理和预测方法, 而人体行为传感器数据又属于时间序列上的信号, 因此本文将 LSTM 层作为所提 CLT-net 网络模型的特征筛选单元引入。另外, 作为对比构建的长短时记忆网络模型如图 2 所示, 主要包括序列输入层、flatten 层、LSTM 层、全连接层、softmax 分类层。

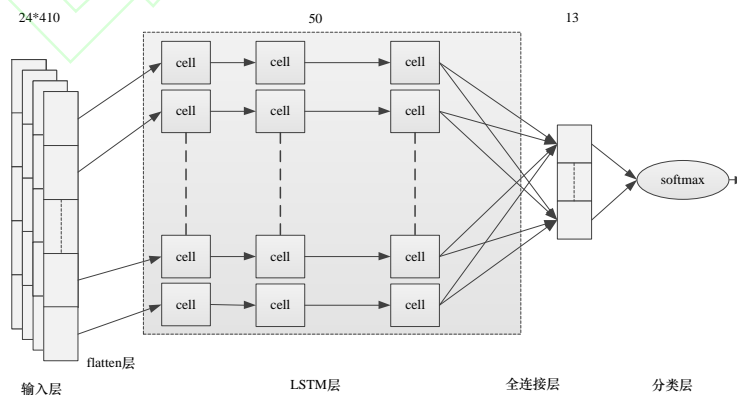


图 2 LSTM 模型结构

Fig.2 LSTM model structure

图 2 中, 序列输入层样本大小为 $24 \times 410 \times 1$, 经过 flatten 层将多维数据一维化后作为 LSTM 层的输入, LSTM 层的隐藏单元数量设置为 50, 全连接层隐藏节点设置为 13, 最终由 softmax 分类层实现对不同人体行为的分类。

LSTM 层的细胞(cell)为输入数据提供了时间依赖性, 赋予了数据时间特征, LSTM 网络通过 cell 实现长期控制, 进而用于时序信号的分类预测。细胞 (cell) 功能主要是通过遗忘门、输入门和输出门实现的。LSTM 层 cell 的内部结构如图 3 所示。

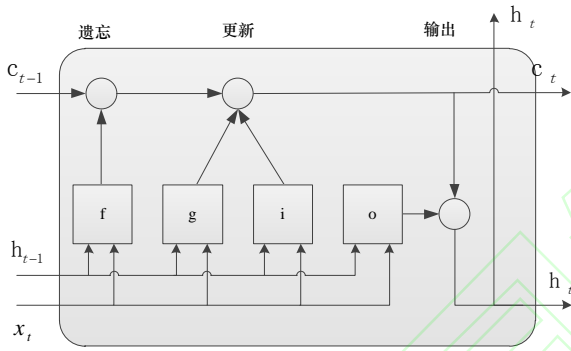


图 3 LSTM 细胞内部结构

Fig.3 The internal structure of LSTM cell

LSTM 层可学习权值为输入权重 W (Input Weights)、递归权重 R (Recurrent Weights) 和偏差 b (Bias)。矩阵 W 、 R 和 b 分别是输入权重、递归权重和每个分量的偏差的串联。这些矩阵连接如下：

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \quad (3)$$

图 3 中, i 、 f 、 o 、 g 分别为输入门、遗忘门、输出门和候选输入单元。 t 时刻细胞状态输出和隐藏状态输出表达式分别为：

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (4)$$

$$h_t = o_t \odot \sigma_c(c_t) \quad (5)$$

其中, \odot 为 Hadamard 乘积 (向量的元素相乘), σ_c 为双曲正切函数 (\tanh) 状态激活函数。

图 3 中, t 时刻遗忘激活 f_t 、输入激活 i_t 、输出激活 o_t 、候选单元输入 g_t 的表达式为：

$$f_t = \sigma_g(W_f x_t + R_f h_{t-1} + b_f) \quad (6)$$

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i) \quad (7)$$

$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o) \quad (8)$$

$$g_t = \sigma_c(W_g x_t + R_g h_{t-1} + b_g) \quad (9)$$

式中, h_{t-1} 为上一时刻隐藏状态的输出信息, x_t 为当前时刻的输入信息, 即将 h_{t-1} 与 x_t 共同作为当前时间步的输入信息参与网络训练。这些信息经过门激活函数 σ_g (sigmoid) 后, 最终得到输出介于 $[0,1]$ 之间的值。

且遗忘激活 f_t 越大, 代表遗忘上一时刻细胞状态输出 c_{t-1} 越少; 输入激活 i_t 越大, 代表候选输入 g_t 被写进当前时刻的信息越多, 即遗忘激活 f_t 和输入激活 i_t 共同决定了当前时刻细胞状态输出 c_t 对不同输入信息的接收程度; 输出激活 o_t 则确定了当前时刻隐藏状态的输出 h_t , 以上的控制策略便实现了人体行为数据在时间步长序列上的长期依赖性。

2.3 CLT-net 人体行为识别模型

由于惯性传感器采集的人体行为数据可以看作是时间序列信号, 以及当前机器学习算法严重依赖手工设计的特征, 可能会导致信息利用不充分、无法有效实现对复杂人体活动的识别等问题, 本文提出了基于空时特征融合的深度学习模型 (CLT-net) 用于人体行为识别。CLT-net 模型充分结合了 CNN 可

以自动提取数据深层特征和 LSTM 能够学习时序数据之间的相关性的优点,采用与上述 CNN 模型相同的网络结构,同时使用 LSTM 层替换了 CNN 模型的第一个全连接层,CLT-net 网络模型各项参数的设

置和规范函数的选择与对应结构的 CNN 模块和 LSTM 模块一致,其结构如图 4 所示,主要包括序列输入层、折叠层、CNN 特征提取层、解折叠层、flatten 层、LSTM 层、全连接层、softmax 分类层。

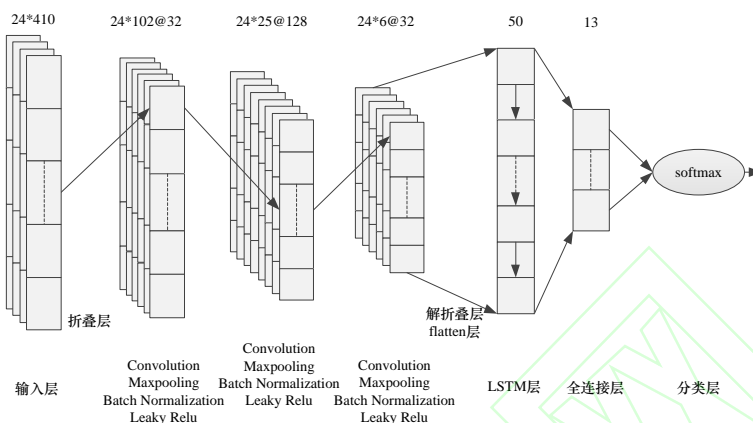


图 4 CLT-net 模型结构

Fig.4 CLT-net model structure

基于 CLT-net 模型的人体行为识别方法的分类过程大致如下:首先,将输入的人体行为数据序列经过 CNN 模块进行二维空间上的特征提取;然后,将得到的二维数据特征经过 flatten 层展开成一维后输入 LSTM 层进行时间序列上的特征筛选;其次,再经过全连接层将筛选后得到的人体行为特征通过权重矩阵映射至样本标记空间;最后通过 softmax 层进行分类计算,并选择具有最大预测概率的类别作为输入数据样本的预测类别。模型在训练时会根据前向传播得到的预测类别与真实样本标签之间的误差,并依据模型采用的损失函数和优化器进行反向传播以不断地修正网络中的权值和偏置项,最终实现模型训练并得到较优模型。

3 实验设置

3.1 实验数据集介绍

本文基于可穿戴传感器数据进行人体行为识别研究,实验采用公开的 DaLiAc (Daily Life Activities)

数据集^[25]。该数据集的数据采集过程由分别放置在受试者右臀部、胸部、右手腕和左脚踝 4 个 6 轴惯性传感器节点完成,每个传感器节点由一个三轴加速度计和一个三轴陀螺仪组成。加速度计的量程为 $\pm 6\text{ g}$,手腕、胸部、臀部传感器节点的陀螺仪范围为 $\pm 500\text{ deg/s}$,踝关节传感器节点的陀螺仪范围为 $\pm 2000\text{ deg/s}$,数据的采样频率为 204.8 Hz 。数据采集实验共有 19 名健康受试者参与(女性 8 名,男性 11 名,年龄 (26 ± 8) 岁,身高 $(177\pm 11)\text{ cm}$,体重 $(75.2\pm 14.2)\text{ kg}$,偏差 $\text{mean} \pm \text{std}$),共采集了 13 项活动。活动及对应标签如表 1 所列。

表 1 活动及对应标签

Table1 Activities and corresponding labels

Activities	活动描述	labels
Sitting	端坐	1
Lying	平躺	2
Standing	站立	3
Washing dishes	洗盘子	4
Vacuuming	吸尘	5
Sweeping	扫地	6
Walking	步行	7
Ascending stairs	上楼	8
Descending stairs	下楼	9
Treadmill running	跑步机跑步	10
Bicycling on ergometer (50W)	测功单车骑车 (50W)	11
Bicycling on ergometer (100W)	测功单车骑车 (100W)	12
Rope jumping	跳绳	13

3.2 实验运行环境

本文中的所有模型均在配置为 Core i5-6500U CPU @3.20GHz, 16GB 内存的计算机上进行了训练

和测试, 该计算机系统为 Windows10 专业版 64 位, 并且所有模型是使用 Matlab2020b Deep Learning Toolbox 框架实现。

3.3 实验相关参数

实验时首先对人体行为数据进行样本的划分, 取滑动窗口的长度为 410 (2 倍的采样频率取整), 且相邻窗口之间存在 50% 的数据重叠, 因此单个样本序列大小为 (24×410) (4 个 6 轴传感器)。将经过数据分割后的样本打乱顺序, 取前 90% 的样本作为训练集, 剩下 10% 的数据作为测试集。实验参数的选择如表 1 所列, 仿真实验时所有模型的初始化参数都使用相同配置, 这样做的目的是使得所有模型都在相对公平的条件进行对比, 从而更有利于反映出 CNN 模型、LSTM 模型和 CLT-net 模型的真实性能。

表 2 实验相关参数

Table2 Experimental parameters

CNN 层权系数初始化	LSTM 层权系数初始化	全连接层权系数初始化	优化器	损失函数	初始学习率
Kaiming 法	正交法	Kaiming 法	Adam 优化器	交叉熵	0.001
样本序列大小	训练集样本数	测试集样本数	训练轮数	批尺寸	Leaky Relu 因子
24×410	20088	2232	20	500	0.1

其中, CNN 层和全连接层的权系数初始化采用 Kaiming 法, 该方法有利于加快模型的收敛速度, LSTM 层的权系数初始化方法采用正交方法。所有模型的优化器都使用自适应矩估计优化算法(Adam 算法)^[26], 该方法具有更快的收敛速度和更低的内存消耗需求, 并且在模型训练过程中可以不使用验证集。

4 实验结果与分析

4.1 模型收敛速度

LSTM 模型、CNN 模型和 CLT-net 是在训练集上进行训练的, 模型训练过程的准确率曲线和损失

函数曲线如图 5 和图 6 所示。

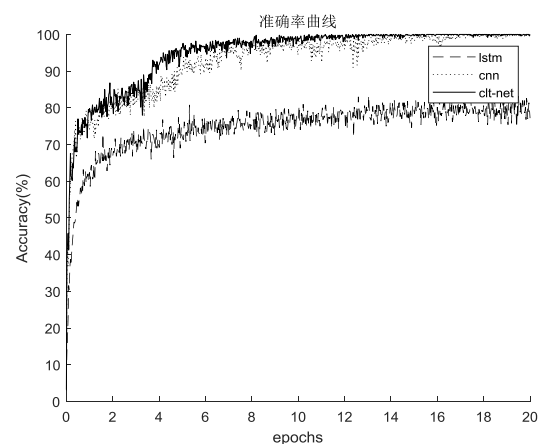


图 5 LSTM、CNN 和 CLT-net 训练过程准确率曲线

Fig.5 Accuracy curves of LSTM, CNN and CLT-net in training process

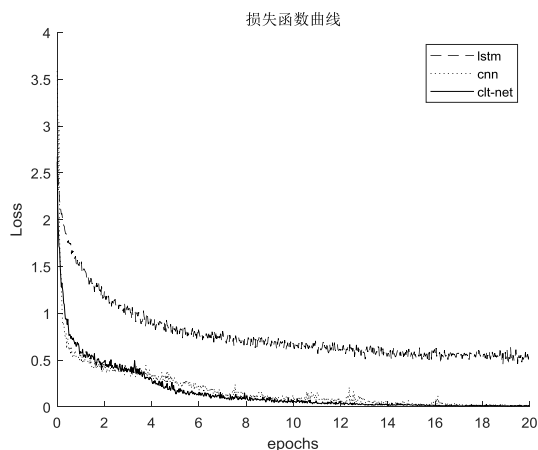


图 6 LSTM、CNN 和 CLT-net 训练过程损失函数曲线

Fig.6 Loss function curves of LSTM, CNN and CLT-net in training process

图 5 中三种模型训练过程的准确率曲线随着迭代次数 (epochs) 的增加逐渐趋近于 100%，说明了模型对不同人体活动的分类结果越来越好。图 6 中三种模型训练过程的损失函数曲线随着迭代次数 (epochs) 的增加逐渐趋近于 0，说明了模型各项参数的修正更新逐渐向较优值靠近。

综合图 5 和图 6 也可以看出，CLT-net 模型具有最快的收敛速度使得训练集的分类准确率和损失函数值接近稳定状态，且随着迭代次数的增加 CLT-net 模型的分类准确率逐渐上升至最高、损失值逐渐降低至最小，而 CNN 和 LSTM 则分别次之，证明了 CLT-net 模型用于人体活动识别的高效性。

4.2 模型准确性

图 7、图 8 和图 9 分别对应训练好的 LSTM 模型、CNN 模型和 CLT-net 以及 BP 神经网络算法在测试集上的预测分类混淆矩阵，以及作为对比建立的 BP 神经网络模型在测试集上的预测分类混淆矩阵如图 10 所示。

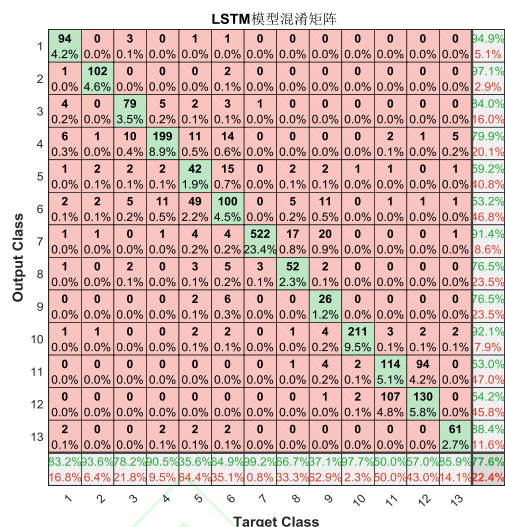


图 7 LSTM 模型测试集的预测分类混淆矩阵

Fig.7 Prediction classification confusion matrix of LSTM model test set

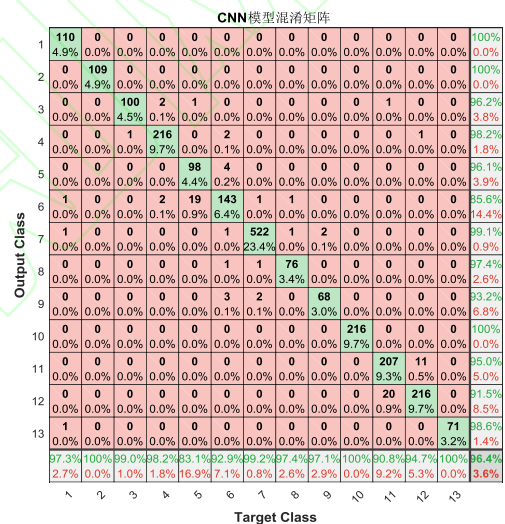


图 8 CNN 模型测试集的预测分类混淆矩阵

Fig.8 Prediction classification confusion matrix of CNN model test set

CLT-net模型混淆矩阵													
Output Class	1	2	3	4	5	6	7	8	9	10	11	12	13
1	110 4.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	109 4.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	99 4.4%	2 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.0% 2.0%
4	2 0.1%	2 0.0%	217 9.7%	0 0.0%	4 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	96.0% 4.0%
5	1 0.0%	0 0.0%	0 0.0%	0 0.0%	104 4.7%	3 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	95.4% 4.6%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	14 0.6%	145 6.5%	1 0.0%	2 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	89.0% 11.0%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	523 23.4%	1 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.6% 0.4%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	74 3.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.1%	2 0.1%	69 3.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	94.5% 5.5%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	216 9.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
11	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	218 9.8%	3 0.1%	0 0.0%	0 0.0%	98.6% 1.4%
12	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	10 0.4%	224 10.0%	0 0.0%	95.7% 4.3%
13	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	71 3.2%	100% 0.0%
Target Class	1	2	3	4	5	6	7	8	9	10	11	12	13
	97.3% 2.7%	100% 0.0%	98.0% 2.0%	98.6% 1.4%	88.1% 11.9%	84.2% 15.8%	89.4% 10.6%	84.9% 15.1%	98.6% 1.4%	100% 0.0%	95.6% 4.4%	98.2% 1.8%	100% 0.0%

图 9 CLT-net 模型测试集的预测分类混淆矩阵

Fig.9 Prediction classification confusion matrix of CLT-net model test set

BP模型混淆矩阵													
Output Class	1	2	3	4	5	6	7	8	9	10	11	12	13
1	14 0.6%	9 0.4%	25 1.1%	3 0.1%	0 0.0%	7 0.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	24.1% 75.9%
2	13 0.6%	8 0.4%	7 0.3%	7 0.3%	0 0.0%	2 0.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	21.6% 78.4%
3	54 2.4%	61 2.7%	51 2.3%	32 1.4%	3 0.1%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	25.2% 74.8%
4	5 0.2%	4 0.2%	9 0.4%	159 7.1%	2 0.1%	5 0.2%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	0 0.0%	35.5% 64.5%
5	0 0.0%	2 0.1%	10 0.4%	56 2.5%	49 2.2%	1 0.0%	5 0.2%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	44.4% 55.6%
6	2 0.1%	1 0.0%	1 0.0%	3 0.1%	54 2.4%	78 3.5%	0 0.0%	56 2.5%	6 0.3%	3 0.1%	6 0.3%	0 0.0%	37.1% 62.9%
7	0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	488 21.9%	0 0.0%	5 0.2%	11 0.5%	58 2.6%	4 0.2%	5 0.2%	85.3% 14.7%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.1%	2 0.1%	0 0.0%	2 0.1%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	22.2% 77.8%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	34 1.5%	9 0.4%	52 2.3%	0 0.0%	1 0.0%	0 0.0%	52.0% 48.0%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	158 7.1%	2 0.1%	0 0.0%	2 0.1%	95.8% 4.2%
11	24 1.1%	23 1.0%	6 0.3%	5 0.2%	0 0.0%	3 0.1%	0 0.0%	0 0.0%	0 0.0%	128 5.7%	97 4.3%	0 0.0%	44.8% 55.2%
12	1 0.0%	0 0.0%	0 0.0%	1 0.0%	2 0.1%	2 0.1%	1 0.0%	1 0.0%	93 4.2%	120 5.4%	0 0.0%	0 0.0%	55.0% 45.0%
13	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	63 2.8%	100% 0.0%
Target Class	1	2	3	4	5	6	7	8	9	10	11	12	13
	12.4% 87.6%	7.3% 92.7%	50.5% 49.5%	72.3% 27.7%	47.5% 52.5%	50.6% 49.4%	82.8% 17.2%	2.6% 97.4%	74.3% 25.7%	73.1% 26.9%	56.1% 43.9%	52.6% 47.4%	88.7% 11.3%

图 10 BP 模型测试集的预测分类混淆矩阵

Fig.10 Prediction classification confusion matrix of BP model test set

4 个混淆矩阵的含义如下：分别为 14 行 14 列。

底侧数字 1 到 13 代表待分类的 13 种人体行为，左侧数字 1 到 13 代表预测分类出的 13 种人体行为。最后一行格子（右下角格子除外）上面和下面的百分比分别表示模型对某一人体行为做出正确或错误分类的准确率（即召回率，recall）和错误率。最后一列格子（右下角格子除外）上面和下面的百分比

分别表示模型预测分类为某一人体行为的准确率（即精确率，precision）和错误率。右下角格子上的百分比代表模型对 13 种人体行为做出正确分类的总体平均分类准确率，下面的百分比为总体平均分类错误率。其他格子下面的百分比则表示该分类样本数占全部测试集样本数的比例。

从 4 个混淆矩阵中可以看出，本文提出的 CLT-net 具有最高的总体平均分类准确率，达到了 97.6%，相比较 BP 神经网络、LSTM 和 CNN 模型的 61.7%、77.6%和 96.4%分别有 35.9%、20.0%和 1.2%的提升。仅比较 LSTM 模型和 CNN 模型也可以发现，CNN 模型具有更高的识别率，这是因为 CNN 能够提取人体行为数据的特征，这些特征最大限度的代表了原始人体行为数据，用这些特征进行人体行为识别分类具有较好的表现。而 LSTM 只是用于时序数据的建模以学习数据间的相关性，并不能实现特征的提取，这也说明了特征提取是进行分类识别的关键，CNN 特征提取模块是 CLT-net 模型最重要的组成部分。

4.3 模型分类结果可视化

t-SNE(t-Distribution Stochastic Neighbour Embedding, t 分布随机近邻嵌入) 是一种适合高维数据可视化的降维算法^[27]。t-SNE 核心思想是将高维点嵌入低维点，同时尊重点之间的相似性，高维空间中的附近点对应于附近嵌入的低维点，高维空间中的远处点对应于远处嵌入的低维点，通过可视化低维点以查看原始高维数据中的自然簇。本文利用 t-SNE 高维数据可视化方法，将人体行为的分类结果通过可视化手段更加直观呈现出来。图 10、图 11 和图 12 分别展示了 LSTM 模型、CNN 模型和

CLT-net 在测试集上的人体行为分类结果。

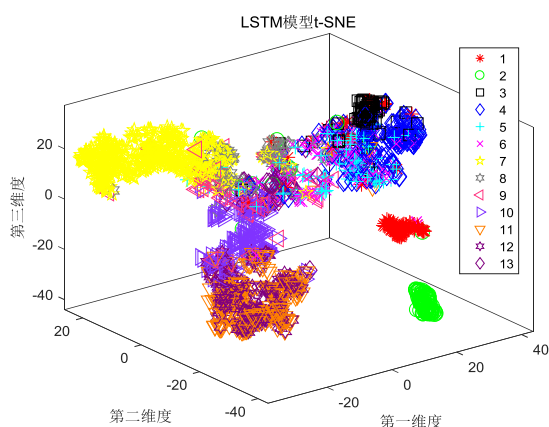


图 11 LSTM 模型的人体行为分类结果可视化

Fig.11 Visualization of human activity classification results based on LSTM model

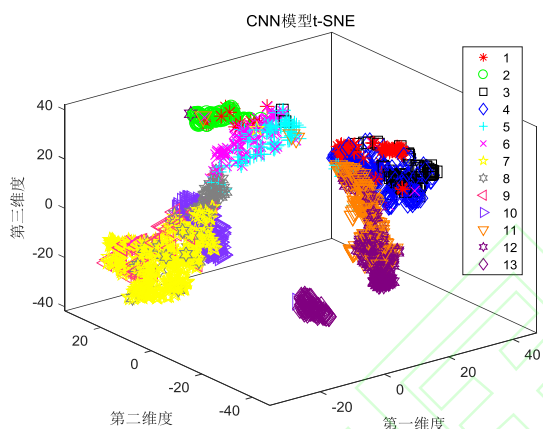


图 12 CNN 模型的人体行为分类结果可视化

Fig.12 Visualization of human activity classification results based on CNN model

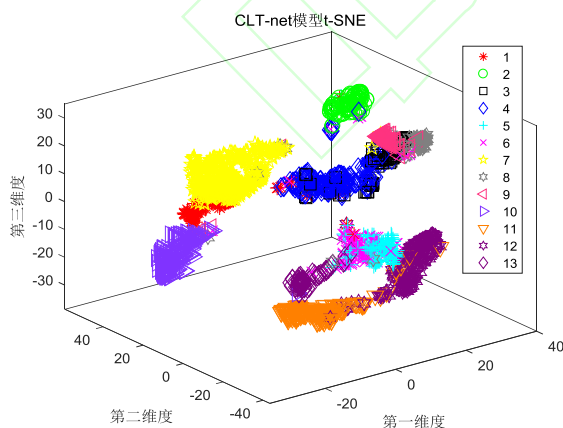


图 13 CLT-net 模型的人体行为分类结果可视化

Fig.13 Visualization of human activity classification results based on CLT-net model

从 3 个模型的 t-SNE 可视化图形中可以明显的

看出, 基于 CLT-net 模型的 t-SNE 数据可视化能够把 13 种人体行为很好的分类 (13 种人体行为分别被聚成一簇), CNN 模型和 LSTM 模型则分别次之, 这也反映了我们所提 CLT-net 模型具有优越的人体行为分类识别性能。

4.4 模型评价指标

为了更好的说明本文所提 CLT-net 模型的泛化能力, 实验进一步统计计算了 LSTM、CNN、CLT-net 和传统 BP 神经网络算法在测试集上测试结果的宏查准率 (macro precision)、宏查全率 (macro recall) 和宏 F1 值 (macro F1-score), 即对 13 种人体行为类别中每一类行为的 precision、recall 和 F1-score 进行求和再平均得到的值^[28], 其结果如表 3 所列。

表 3 四种模型的评价指标对比

Table 3 Comparison of evaluation indexes of four models

	Macro precision	macro recall	macro F1-score
BP	0.5332	0.5238	0.5103
LSTM	0.7696	0.7228	0.7371
CNN	0.9621	0.9614	0.9612
CLT-net	0.9746	0.9716	0.9727

Precision 评价指标是衡量正确预测出的正样本数占实际预测出正样本数的比例, Recall 评价指标是衡量正确预测出的正样本数占总正样本数的比例, 而 F1-score 评价指标作为 precision 和 recall 这两个指标的调和平均值, 其能够更好的对比出 4 种模型对 13 种人体行为的分类性能。从表 3 中可以看出, CLT-net 模型的 precision、recall 和 F1-score 评价指标均最高, 而 BP 模型的所有指标均最低, 充分证明了 CLT-net 模型对不同人体行为的识别分类结果具有良好的稳定性与可靠性。

5 结束语

针对人体行为传感器时序数据的特点, 以及目

前机器学习分类监测识别性能存在的问题,本文提出了基于空时特征融合技术的深度学习模型 CLT-net,并且在 DaLiAc 数据集上 LSTM 模型和 CNN 模型以及传统的 BP 神经网络算法进行了对比实验,通过比较模型训练过程的收敛速度,模型在测试集上的预测分类性能及人体行为分类可视化结果和模型的评价指标,展现了本文提出的模型 CLT-net 在人体行为识别领域的高效性。

目前,在基于传感器的人体行为识别方面的深度学习模型较少,未来可以使用深度学习网络作为人体行为识别特征学习方面进行性能改善的方法,同时将构建的深度学习模型应用到实时人体行为检测。此外,随着边缘计算技术的兴起,打造轻量级的深度学习模型或将成为未来的发展趋势。

参考文献

- [1] Fu B , Damer N , Kirchbuchner F , et al. Sensing Technology for Human Activity Recognition: a Comprehensive Survey[J]. IEEE Access, 2020, 8:83791-83820.
- [2] Hassan M M , Uddin M Z , 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.
- [3] Dang L M , Min K , Wang H , et al. Sensor-based and vision-based human activity recognition: A comprehensive survey[J]. Pattern Recognition,2020,108:107561.
- [4] Tu Z , Xie W , Qin Q , et al. Multi-stream CNN: Learning representations based on human-related regions for action recognition[J]. Pattern Recognition, 2018, 79:32-43.
- [5] Li J , Xie X , Pan Q , et al. SGM-Net: Skeleton-Guided Multimodal Network for Action Recognition[J]. Pattern Recognition, 2020, 104:107356.
- [6] Bulling A , Blanke U , Schiele B . A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors[J]. Acm Computing Surveys, 2014, 46(3):1-33.
- [7] Ferhat A , Samer M , Mariam D , et al. Physical Human Activity Recognition Using Wearable Sensors[J]. Sensors, 2015, 15(12):31314-31338.
- [8] Lima W S , Souto E , El-Khatib K , et al. Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview[J]. Sensors (Basel, Switzerland), 2019, 59:235-244.
- [9] Jobanputra C , Bavishi J , Doshi N. Human Activity Recognition: A Survey[J]. Procedia Computer Science, 2019, 155:698-703.
- [10] Wang J , Chen Y , Hao S , et al. Deep Learning for Sensor-based Activity Recognition: A Survey[J]. Pattern Recognition Letters, 2017,119:3-11.
- [11] 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 Applications, 2018, 105:233-261.
- [12] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [13] Kou Dalei,Quan Jichuan,Zhang Zhongwei.Research on Progress of Object Detection Framework Based on Deep Learning[J]. Computer Engineering and Applications,2019,55(11):25-34. (in Chinese)
寇大磊,权冀川,张仲伟.基于深度学习的目标检测框架进展研究[J].计算机工程与应用,2019,55(11):25-34.
- [14] Tu Z , Li H , Zhang D , et al. Action-Stage Emphasized Spatio-Temporal VLAD for Video Action Recognition[J] . IEEE Transactions on Image Processing, 2019, 28(6):2799-2812.
- [15] Yong D , Yun F , Liang W . Representation Learning of Temporal Dynamics for Skeleton-Based Action Recognition[J]. IEEE Transactions on Image Processing, 2016, 25(7):3010-3022.
- [16] Young T , Hazarika D , Poria S , et al. Recent Trends in Deep Learning Based Natural Language Processing[J]. IEEE Computational Intelligence Magazine , 2018, 13(3):55-75.
- [17] Li X , Yi X , Liu Z , et al. Application of novel hybrid deep leaning model for cleaner production in a paper industrial wastewater treatment system[J]. Journal of Cleaner Production, 2021, 294:126343.

- [18] Abdulmajid M , Jae-Young P . Deep Recurrent Neural Networks for Human Activity Recognition[J]. Sensors, 2017, 17(11):2556.
- [19] Lee S M , Yoon S M , Cho H . Human activity recognition from accelerometer data using Convolutional Neural Network[C]// 2017 IEEE International Conference on Big Data and Smart Computing(BigComp). Washington D. C., USA: IEEE Press, 2017:131-134.
- [20] Ha Q D , Tran M T . Activity Recognition from Inertial Sensors with Convolutional Neural Networks[C]// International Conference on Future Data and Security Engineering. Cham, Germany: Springer, 2017:285-298.
- [21] Fan Changjun, Gao Fei. Human Daily Activity Recognition Based on Wearable sensors[J]. Chinese Journal of Sensors and Actuators, 2018, 31(7):1124-1131. (in Chinese)
范长军, 高飞. 基于可穿戴传感器的普适化人体活动识别[J]. 传感技术学报, 2018, 31(7):1124-1131.
- [22] Aljarrah A A , Ali A H . Human Activity Recognition using PCA and BiLSTM Recurrent Neural Networks [C]// 2019 2nd International Conference on Engineering Technology and its Applications (IICETA). Washington D. C., USA: IEEE Press, 2019: 156-160.
- [23] Hochreiter S , Schmidhuber J . Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [24] Yu Y , Si X , Hu C , et al. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures[J]. Neural Computation, 2019, 31(7):1235-1270.
- [25] Leutheuser H , Schuldhaus D , Eskofier B M. Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset[J]. PLOS ONE, 2013, 8(10): e75196.
- [26] ZHAO Xiaoqiang, SONG Zhaoyang. Adam Optimized CNN Super-Resolution Reconstruction[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(5):858-865. (in Chinese)
赵小强, 宋昭漾. Adam 优化的 CNN 超分辨率重建[J]. 计算机科学与探索, 2019, 13(5):858-865.
- [27] Wei Shichao, Li Xin, Zhang Yichi , et al. Dimension Reduction and Visualization of Mixed-Type Data Based on E-t-SNE [J]. Computer Engineering and Applications, 2020, 56(6):66-72. (in Chinese)
魏世超, 李歆, 张宜弛, 等. 基于 E-t-SNE 的混合属性数据降维可视化方法 [J]. 计算机工程与应用, 2020, 56(6):66-72.
- [28] Chen Minghao, Zhu Yuefei, Lu Bin, et al. Classification of Application Type of Encrypted Traffic Based on Attention-CNN [J]. Computer Science, 2021, 48 (4): 325-332. (in Chinese)
陈明豪, 祝跃飞, 芦斌, 等. 基于 Attention-CNN 的加密流量应用类型识别[J]. 计算机科学, 2021, 48(4):325-332.