

基于协同 LSTM 神经网络的人体行为识别研究

朱连章 陈殿明 郭加树 张红霞

(中国石油大学(华东) 计算机与通信工程学院, 山东 青岛 266580)

摘要: 在基于传感器的人体行为识别研究中,传统的机器学习方法需要具备一定的身体运动领域知识来做特征提取,而且工程量大。而现有的神经网络模型结构简单,对数据特征的挖掘不充分从而识别准确率不高。针对上述问题,提出一种基于协同 LSTM 神经网络的人体行为识别方法。该方法首先对 LSTM 模块的结构进行改进,搭建协同 LSTM 神经网络;然后使用加速度传感器和陀螺仪获取 6 轴人体行为数据;再使用滑窗方法和改进的 Z-score 标准化方法对数据进行预处理;最后利用协同 LSTM 神经网络、卷积神经网络和 LSTM 神经网络分别在数据集上进行迭代训练和测试。实验结果表明,基于协同 LSTM 神经网络的识别模型表现最好,识别准确率为 95.81%,高于 CNN 的 91.53% 和 LSTM 的 90.47%,证明该方法可以有效地进行人体行为识别。

关键词: 人体行为识别; LSTM; 神经网络; 加速度传感器; 陀螺仪

中图分类号: TP389.1

文献标识码: A

文章编号: 1673-629X(2018)12-0079-04

doi: 10.3969/j.issn.1673-629X.2018.12.017

Research on Human Action Recognition Based on Synergistic LSTM Neural Network

ZHU Lian-zhang, CHEN Dian-ming, GUO Jia-shu, ZHANG Hong-xia

(School of Computer & Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: In the sensor-based human action recognition research, feature extraction by traditional machine learning methods needs lots of work and considerable knowledge in the field of human activity, while the existing neural networks' recognition accuracy is not high for its simple structure and insufficient feature mining. For this, we propose a new human action recognition approach based on synergistic LSTM neural network. Firstly the structure of LSTM module is improved and the synergistic LSTM neural network is built; then six-axis human action data are acquired by accelerometer and gyroscope; next, the data will be preprocessed by sliding window and the Z-score normalization; at last, iterative training and testing are performed on the dataset using synergistic LSTM neural network, convolutional neural network and LSTM neural network respectively. The experiment shows that the proposed approach is the best with recognition accuracy by 95.81% higher than 91.53% of CNN and 90.47% of LSTM, which is proved to be effective for human action recognition.

Key words: human action recognition; LSTM; neural network; accelerometer; gyroscope

0 引言

随着科学技术的飞速发展,便携式和穿戴式智能设备逐渐在生活与工作领域中发挥着重要的作用,如人机交互、虚拟现实、运动和医疗保健等^[1]。而通过利用智能设备和计算机技术处理人体行为数据从而实现行为识别成为了国内外的研究热点。

目前人体行为识别的研究主要有两种方式,一是通过录像、拍照的方式获取人体行为的视频、图像数据^[2-3],并对数据进行处理分析。然而该方法较为复

杂,不仅对采集、处理数据的设备要求较高,而且计算量庞大,另一方面可能存在各种无法预料的环境因素对数据处理造成不同程度的干扰,并且在隐私保护、便携性等方面也表现出不足;另一种通过单一或多种传感器获得多维动作数据^[4],再对数据进行处理分析从而实现行为识别。

随着智能终端设备领域的不断发展与进步,在智能终端上集成丰富的传感设备已经非常方便,例如加速度传感器、磁力计、陀螺仪、全球定位系统等已经可

收稿日期: 2018-01-21

修回日期: 2018-05-23

网络出版时间: 2018-09-21

基金项目: 山东省自然科学基金(ZR2014FQ018)

作者简介: 朱连章(1965-),男,博士,教授,硕导,研究方向为软件性能评价、软件测试、油田数据模型;陈殿明(1993-),男,硕士研究生,研究方向为人体行为识别。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20180920.1536.032.html>

以集成在如智能手机、智能手环等可便携、穿戴的设备上,这样就为智能终端设备应用于行为识别提供了可行性^[5-6]。

Ling 等^[7]用 5 个小双轴加速度传感器同时佩戴在四肢和右髋部位来收集个体日常动作数据,并比较不同分类器后发现使用决策树能够获得最佳性能,识别准确率为 84%。Tapia 等^[8]利用心率监视器以及附在四肢及腰部的五个三轴加速度计来组成识别系统,对三十个体育项目动作进行识别,获得了 80.6% 的识别准确率。文献[9-10]中使用了神经网络与深度学习,虽然特征提取工程量小,但是网络结构简单,且识别准确率不足。李锋等^[11]使用单一的加速度传感器进行识别,虽然识别准确率达到 96.13%,但特征提取工程量太大,需要专业的运动领域知识,实用性不足。

针对上述问题,提出了一种改进的基于协同长短期记忆模块的神经网络,并基于该神经网络构建了人体行为识别模型,并通过实验对其进行验证。

1 LSTM 神经网络

1.1 递归神经网络 RNN

RNN 是包含循环的网络,在 RNN 网络结构中,隐层节点读取输入层的信息,然后在输出信息的同时,通过循环结构将信息输入下一步从而实现信息的传递。RNN 的链式结构与时间序列类的数据结构契合,是一种适合处理该类数据的神经网络结构。RNN 在语音识别、语言建模等领域都取得了一定的成就。但是随着 RNN 模块之间的间距增加,RNN 会很难达到长范围的依赖,从而出现梯度消失^[12]问题。

1.2 长短期记忆网络

长短期记忆(long short term memory, LSTM)网络由 Hochreiter & Schmidhuber 在 1997 提出^[13]。Alex Graves 近期对其进行了改良和推广,将 LSTM 应用在很多领域,并取得了相当大的成功^[14-16]。

RNN 模块结构简单,比如有的结构中只有一个单一的 tanh 层,而 LSTM 通过刻意的设计来避免梯度消失问题。

图 1 是标准 LSTM 模块的内部结构。

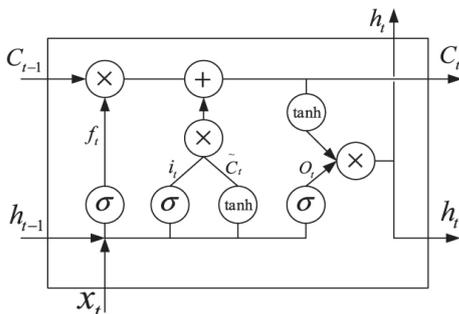


图 1 单个 LSTM 模块结构

LSTM 的关键就在细胞状态 $C_{t-1} \rightarrow C_t$,细胞状态类与传送带相似,直接在整个链上运行,只有少量的线性信息交互,使得细胞信息便于保持。

LSTM 通过特别设计的一种被称为“门”(gates)的结构来拥有增加或去除细胞状态上的信息的能力。

遗忘门的原理如式 1 所示:

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

输入门的原理如式 2 所示:

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \end{cases} \tag{2}$$

细胞更新状态的方式如式 3 所示:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3}$$

输出门原理如式 4 所示:

$$\begin{cases} o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases} \tag{4}$$

2 改进的 LSTM 网络及人体行为识别模型

Gers & Schmidhuber 提出了基于 peephole connection 变体结构的 LSTM^[17],该结构使各门层也接受细胞状态的输入,实验结果表示基于这种变体结构的神经网络比普通的 LSTM 网络表现更好。杨年峰研究了步态特征及其影响因素,并为量化描述人体运动协调规律提供了有效手段^[18]。受 peephole connection 结构与人体运动协调规律的启发,文中对 LSTM 模块的结构进行了改进。

2.1 改进的 LSTM 模块

设计了协同 LSTM 模块(synergistical LSTM, S-LSTM),新的结构中对于更新细胞状态信息时,遗忘门与输入门以同步互补的方式替换标准结构中的分离式更新方法,更改后的细胞状态信息更新方法如式 5 所示:

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t \tag{5}$$

S-LSTM 模块的内部结构如图 2 所示。

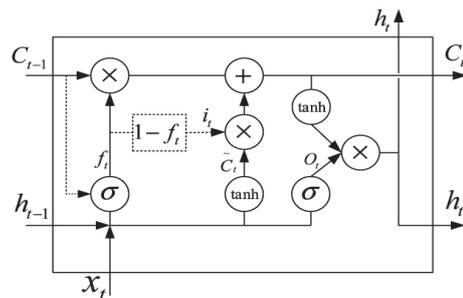


图 2 S-LSTM 模块结构

模块中将 peephole connection 结构应用在细胞历史状态与遗忘门之间,由于式 5 中的更改方式使得遗忘门与输入门同步互补更新,记忆细胞状态在输出部

分子式 4 中已有体现,故而输入门与输出门部分不再添加 peephole connection 结构。此时遗忘门与输入门的更新方式如式 6 所示:

$$\begin{cases} f_i = \sigma(W_f [C_{i-1}, h_{i-1}, x_i] + b_f) \\ i_i = 1 - f_i \end{cases} \quad (6)$$

2.2 基于 S-LSTM 模块的神经网络

文中所用的 S-LSTM 神经网络由以下部分构成: 输入层、6 个 S-LSTM 层、Softmax 分类层和输出层。

2.3 基于 S-LSTM 网络的人体行为识别模型

文中构建了基于 S-LSTM 网络的人体行为识别模型,如图 3 所示。

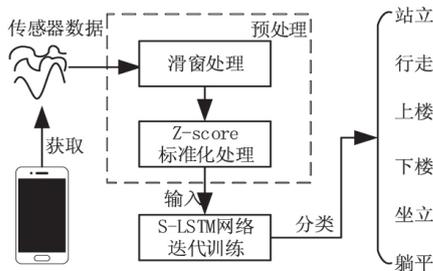


图 3 基于 S-LSTM 网络的人体行为识别模型

2.4 数据预处理方法

使用调整后的 Z-score 标准化方法对数据进行预处理, Z-score 标准化公式如式 7 所示:

$$x' = \frac{x - \mu}{\sigma} \quad (7)$$

其中, μ 为均值; σ 为标准差。

文中对式 7 做了一定的变更,如式 8 所示:

$$x' = \frac{x - \mu}{\sigma + \varepsilon} \quad (8)$$

为避免特定动作使某一轴数据固定从而导致式 7 出现除零错误,引入极小量 ε 。

3 实验

3.1 数据采集

使用的数据集来自公开的 UCI 机器学习知识库 (smartphone-based recognition of human activities and postural transitions data set)^[19],实验数据由 30 名年龄在 19~48 岁的志愿者完成,采集设备为 Samsung Galaxy S II。通过使用加速度传感器与陀螺仪以 50 Hz 的频率获取加速度与角速度数据。志愿者在实验中执行了由 6 项行为动作组成的活动: 3 项静态活动(站立、坐立和躺平)和 3 项动态活动(行走、上楼和下楼)。传感器信号数据通过噪声滤波处理,在 50% 重叠的固定宽度的滑动窗口(2.56 s)中采样,然后使用 Butterworth 低通滤波器分离身体加速度和重力,最终得到 10 929 个样本。

3.2 实验环境和参数设置

实验基于 Windows 10 专业版(版本号 1607),处理

器为 Intel Core i7(2.5 GHz),内存为 8 G,使用支持训练神经网络的 NVIDIA 显卡来提升训练速度。为了验证提出模型的有效性,选择深度卷积神经网络 CNN 和标准 LSTM 神经网络在相同的数据集上进行实验对比。

实验所用数据集被分为两部分:训练集占 70%,测试集占 30%。将样本标签数据做 one-hot 处理,使其与样本数据对应。鉴于显存容量有限,使用 mini-batches 方法来进行批梯度下降。由于已经做了标准化处理(见式 8),L2 损失函数相比于 L1 损失函数表现更佳,所以实验中均使用 L2,优化器使用 Adam,各模型的学习率都设为 0.002,迭代次数为 1 000。

3.3 实验结果分析

三种网络模型在实验中随着迭代次数增加,不断优化参数,并在数据集上进行识别准确率的验证对比。各模型在训练集和测试集上的表现分别如图 4 和图 5 所示。

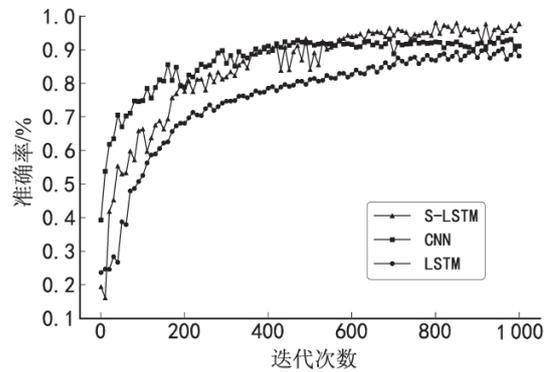


图 4 各模型在训练过程中的识别准确率

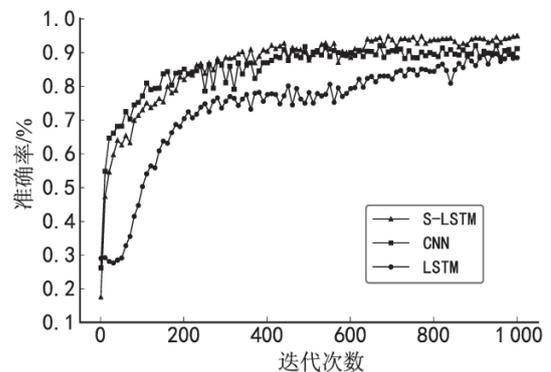


图 5 各模型在测试过程中的识别准确率

可以看到,随着迭代次数的增加,各模型的准确率不断上升,刚开始 CNN 的收敛速度最快,S-LSTM 次之,LSTM 最慢,各模型在测试集上的识别准确率要低于在训练集上的表现。

因为实验本身在一定程度上受参数随机初始化效果的影响,所以对各模型在测试集上都进行 5 次实验,然后取其平均准确率作为对比依据,其对比结果如表 1 所示。

表 1 各模型在测试集上的平均识别准确率 %

神经网络模型	平均识别准确率
CNN	91.53
S-LSTM	95.81
LSTM	90.47

由表 1 可知,三种模型中,S-LSTM 表现最好,准确率达到 95.81%,而 CNN 表现平稳,识别准确率为 91.53%,而标准 LSTM 在迭代过程中的识别准确率波动较大,但最终与 CNN 相差较小,为 90.47%。实验结果表明,提出的基于 S-LSTM 神经网络的人体行为识别模型表现最好,是一种有效的人体行为识别方法。

4 结束语

基于 S-LSTM 神经网络的模型实现了人体行为识别,通过与 CNN、标准 LSTM 神经网络模型在 UCI 人体行为识别数据集上进行了实验对比,表明该方法识别效果最好,说明该方法适合处理人体行为时域数据,能够充分挖掘其数据特征,提高识别准确率。实验使用的是加速度传感器与陀螺仪获取的 6 轴传感数据,目前随着科技的发展,更多的传感器可以集成在便携智能终端设备上,比如磁力计、GPS、摄像头和麦克风等,可以大大丰富数据的多维性。相信在未来的工作中,人们可以通过使用更优秀的网络模型和方法处理基于多传感器的人体行为数据,从而更好地提升识别准确率。

参考文献:

- [1] 谢凌钦,石萍,蔡文杰.可穿戴式智能设备关键技术及发展趋势[J].生物医学工程与临床,2015,19(6):635-640.
- [2] 阮涛涛,姚明海,瞿心昱,等.基于视觉的人体运动分析综述[J].计算机系统应用,2011,20(2):245-254.
- [3] 吴冬梅,谢金壮,王静.基于多特征融合的人体行为识别[J].计算机应用与软件,2015,32(11):171-175.
- [4] CHEN Liming,HOEY J,NUGENT C D,et al. Sensor-based activity recognition[J]. IEEE Transactions on Systems,Man, and Cybernetics,2012,42(6):790-808.
- [5] CHEN Liming,NUGENT C D,WANG Hui. A knowledge-driven approach to activity recognition in smart homes[J]. IEEE Transactions on Knowledge & Data Engineering,2012,24(6):961-974.
- [6] KRISHNAN N C,COOK D J. Activity recognition on streaming sensor data[J]. Pervasive & Mobile Computing,2014,10:138

- 154.
- [7] LING Bao,INTILLE S S. Activity recognition from user-annotated acceleration data[C]//Second international conference on pervasive computing. [s. l.]: [s. n.] 2004:1-17.
- [8] TAPIA E M,INTILLE S S,HASKELL W,et al. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor[C]//11th IEEE international symposium on wearable computers. Boston, MA,USA:IEEE,2007:37-40.
- [9] KWON Y,KANG K,BAE C. Analysis and evaluation of smart-phone-based human activity recognition using a neural network approach[C]//International joint conference on neural networks. Killarney,Ireland:IEEE,2015:1-5.
- [10] ZHANG Licheng,WU Xihong,LUO Dingsheng. Recognizing human activities from raw accelerometer data using deep neural networks[C]//International conference on machine learning and applications. Miami,FL,USA:IEEE,2015:865-870.
- [11] 李锋,潘敬奎.基于三轴加速度传感器的人体运动识别[J].计算机研究与发展,2016,53(3):621-631.
- [12] BENGIO Y,SIMARD P,FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks,2002,5(2):157-166.
- [13] HOCHREITER S,SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.
- [14] 左玲云.基于 LSTM 结构的神经网络语言模型在电话交谈语音识别系统中的应用[D].北京:中国科学院大学,2016.
- [15] 梁军,柴玉梅,原慧斌,等.基于极性转移和 LSTM 递归网络的情感分析[J].中文信息学报,2015,29(5):152-160.
- [16] GRAVES A,ECK D,BERINGER N,et al. Biologically plausible speech recognition with LSTM neural nets[C]//International workshop on biologically inspired approaches to advanced information technology. [s. l.]: [s. n.],2004:127-136.
- [17] GERS F A,SCHMIDHUBER J. Recurrent nets that time and count[C]//IEEE-INNS-ENNS international joint conference on neural networks. Como,Italy:IEEE,2000:3189.
- [18] 杨年峰.人体运动协调规律及其参数化描述[D].北京:清华大学,2001.
- [19] ANGUITA D,GHIO A,ONETO L,et al. A public domain dataset for human activity recognition using smartphones[C]//European symposium on artificial neural networks. Bruges [s. n.] 2013:437-442.