



计算机科学

COMPUTER SCIENCE

基于元学习和注意力机制的用户行为识别研究

王佳昊, 钟鑫, 李文雄, 赵德鑫

引用本文

王佳昊, 钟鑫, 李文雄, 赵德鑫. 基于元学习和注意力机制的用户行为识别研究[J]. 计算机科学, 2023, 50(8): 193-201.

WANG Jiahao, ZHONG Xin, LI Wenxiong, ZHAO Dexin. [Human Activity Recognition with Meta-learning and Attention](#) [J]. Computer Science, 2023, 50(8): 193-201.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于字符特征的 DGA 域名检测方法研究综述](#)

Survey of DGA Domain Name Detection Based on Character Feature

计算机科学, 2023, 50(8): 251-259. <https://doi.org/10.11896/jsjcx.220700277>

[融合音字特征转换的非自回归Transformer中文语音识别](#)

Non-autoregressive Transformer Chinese Speech Recognition Incorporating Pronunciation-Character Representation Conversion

计算机科学, 2023, 50(8): 111-117. <https://doi.org/10.11896/jsjcx.220600144>

[基于注意力机制的多模态在线评论有用性预测研究](#)

Study on Multimodal Online Reviews Helpfulness Prediction Based on Attention Mechanism

计算机科学, 2023, 50(8): 37-44. <https://doi.org/10.11896/jsjcx.220600204>

[语义风格一致的任意图像风格迁移](#)

Arbitrary Image Style Transfer with Consistent Semantic Style

计算机科学, 2023, 50(7): 129-136. <https://doi.org/10.11896/jsjcx.220700008>

[基于变分持续贝叶斯元学习的推荐算法](#)

Variational Continuous Bayesian Meta-learning Based Algorithm for Recommendation

计算机科学, 2023, 50(7): 66-71. <https://doi.org/10.11896/jsjcx.220900125>

基于元学习和注意力机制的用户行为识别研究

王佳昊¹ 钟鑫¹ 李文雄¹ 赵德鑫²

1 电子科技大学信息与软件工程学院 成都 610051

2 军事科学院国防科技创新研究院 北京 100071

(wangjh@uestc.edu.cn)

摘要 深度学习技术研究的深入,极大地促进了其在行为识别领域的应用和发展。目前基于深度学习的行为识别研究不可避免地依赖于大量的训练数据,而基于传感器数据的行为识别问题往往在实际应用中需要面向不同的新用户,导致存在用户数据个性化的问题且难以解决,即不同个体在进行同一行为动作时不可避免地会产生一些数据差异,模型在面对新用户时并不能保证对其具有良好的预期行为识别度,而每次针对新用户采集大量训练数据以进行重新建模缺乏实施可行性。针对这一难题,小样本学习技术在新的任务上仅使用少量数据就能够达到较好效果,即在行为识别问题上,每个新用户仅需采样少量的数据即可完成分类。文中结合小样本学习和行为识别算法,提出了新的解决方案——MAML-M模型。首先采用基于优化的元学习方法根据用户类型对数据集进行划分,并将其构建为多个用户任务用于训练和测试;然后在MAML-M模型中引入了MAML方法以及基于注意力机制的Memory模块;最后提高模型网络提取并归纳数据特征的能力。通过在MEX数据集上进行对比实验,结果证明在小样本设定下,所提出的MAML-M模型优于传统的深度学习方法。

关键词: 人体行为识别;小样本学习;元学习;注意力机制

中图法分类号 TP181

Human Activity Recognition with Meta-learning and Attention

WANG Jiahao¹, ZHONG Xin¹, LI Wenxiong¹ and ZHAO Dexin²

1 School of Information and Software Engineering, University of Electronic Science and Technology, Chengdu 610051, China

2 National Innovation Institute of Defense Technology, Academy of Military Sciences, Beijing 100071, China

Abstract With the in-depth research of deep learning technology, its application and development in the field of behavior recognition have been greatly promoted. Current research on behavior recognition based on deep learning usually requires a large training data set. But when facing practical applications, new users will inevitably run into personalization issues. This means that even while performing the same activity, different people may use training data sets differently. Existing solutions cannot guarantee to achieve the expected accuracy when dealing with new users. Besides, these models would also be impractical to deploy when gathering training data for new users. Facing this problem, small-sample learning can achieve better results by using only a small number of samples. This means that in the behavior recognition problem, each new user can be classified using a little training data. In this paper, a MAML-M model is proposed by combining few-shot learning and behavior recognition algorithms. Firstly, an optimization-based meta-learning method is adopted to divide the dataset according to users and construct multiple user tasks for training and testing. Meanwhile, the MAML method and the memory module based on the attention mechanism are introduced into the MAML-M model, which finally improves the ability of the model network to extract and summarize data features. Through experiment on MEX dataset, the proposed MAML-M model shows better performances under small sample sets.

Keywords Human behavior recognition, Small sample learning, Meta-learning, Attention mechanism

1 引言

近年来,人类活动识别(Human Activity Recognition, HAR)技术研究因其在人体保健、步态识别、智慧医疗、运动跟踪和智能家居等应用领域的潜力而受到越来越多的关注并

取得了长足的发展^[1-3]。人类活动识别技术通常基于视频图片以及传感器数据来实现行为动作的识别和分类,而可穿戴式传感器设备,如智能手环、手表等,具有造价低、便于佩戴以及计算开销小等优点,具有广泛的市场普及度和应用基础^[4-6]。

到稿日期:2022-09-14 返修日期:2022-12-08

基金项目:电子科技大学-智小金-智能家居联合研究中心项目(H04W210180);内江市科技孵化和成果转化专项资金(2021KJFH004);四川省科技计划项目(2022YFG0212)

This work was supported by the UESTC-ZHIXIAOJING Joint Research Center of Smart Home(H04W210180), Neijiang Technology Incubation and Transformation Funds(2021KJFH004) and Science and Technology Program of Sichuan Province, China(2022YFG0212).

通信作者:赵德鑫(zhaodx2008@163.com)

同时,深度学习(Deep Learning, DL)方法因其在各个领域的出色表现而获得了人们的认可和青睐,因此近年来许多基于深度学习的 HAR 成果被提出。然而深度学习方法通常需要建立在足够的训练时间和充足的样本数据的基础上,但构建大规模有标记的人类活动数据集需要大量的时间和人力成本;另外个体之间行为差异所带来的样本数据差异化问题也给模型训练带来了极大的困难。因此,有必要找到一种方法来缓解深度学习对训练数据集的依赖。小样本学习(Few Shot Learning, FSL)是一种迁移学习技术,旨在通过已有的、数据量较为丰富的数据集来学习知识并更新模型,归纳并重用相关类(源域)上现有模型的知识,在新的数据量较少但性质相似的数据集或任务(目的域)上,只需少量标记的训练样本,模型就可以快速地达到较好的效果^[7-8]。其主要解决的问题是在新的任务中只使用少量样本即可获得较好的专业模型,这使得训练深度学习模型变得更加容易。

小样本学习在许多行为识别应用中具有广阔的潜力。例如,在安装了某些医疗保健系统后,系统需要具备重新训练的能力,因为以前系统中使用的训练数据可能无法代表新环境(例如新的活动类型、不同的人等);此外,在训练时无法获取足够多的高质量数据,因为用户通常很难提供大量数据用于训练,尤其是针对老年人或残疾人等的特殊服务场景。

本文主要研究使用传感器数据进行行为识别的方法。值得一提的是,传感器在某些活动领域比其他领域更能捕捉到个人特征和细微差别,尤其在涉及更大自由度的活动时,这种“个性化”差别极为明显,甚至不同的个体执行同一种行为动作所产生的数据也会因为这种“干扰”而产生较大的差异,从而被模型识别为其他种类的行为动作。因此,使用 HAR 数据集识别活动类别的传统的模型并不能很好地适应个性化需求,而更好的个性化意味着更多的同源数据(如对单个用户采集大量数据),这在现实场景中存在极大挑战。结合小样本学习的理念,可以将每个人的全部数据视为一个数据集或作为一个训练任务,但在每个类中仅使用有限数量的数据实例进行模型训练,可以被看作是一个小样本分类任务^[9]。基于元学习(Meta Learning)方法的小样本学习注重“学会学习”的能力,可以从训练任务中对知识进行归纳总结,从而更加快速地适应新的任务。新任务与旧任务并不相同,但要求具备一定的相似性,正是由于这种相似性,小样本学习模型才能更好地利用其在训练阶段学习到的先验知识^[10]。而模型无关的元学习算法(Model Agnostic Meta Learning, MAML)通过学习通用模型初始化参数来实现小样本分类任务,并且只需很少的数据实例就可以快速适应新任务^[11]。“学会学习”的概念与个性化非常吻合,其中建模一个人的行为识别模型可以被视为一项独立任务,因此元学习方法在行为识别任务上的目标是学习一个快速适应新人的行为识别模型。

本文提出了一种基于元学习和注意力的用户行为识别算法,主要贡献如下:

(1)将元学习的思想和方法应用到行为识别领域,结合根据用户划分任务的概念,将独立用户的全部数据作为一个训练任务,使用元学习方法训练若干个用户任务,学习先验知识并进行归纳总结,得到泛化能力较高的初始化参数,使模型仅使用新用户的少量数据便能快速更新到较高精度。

(2)在 MAML-CovLSTM 模型的基础上加入 Memory

模块,提出了 MAML-M 框架,提高了模型的分类能力。

2 相关工作

在以往的基于传感器数据的 HAR 研究中,一种方法是将传感器数据根据时间步长进行拼接组合,变成二维数据,由于其局部依赖性和尺度不变性,卷积神经网络^[12]能够达到优于其他传统的机器学习方法的效果^[13]。对 1DCNN 的研究表明,这些深度学习模型在解决行为识别问题方面比传统的机器学习模型更有效^[14]。同时传感器数据呈现出时间序列的特点,具有一定的时间依赖性,可将其看作时间序列分类问题,使用 LSTM(Long Short Term Memory)网络来进行处理。LSTM 网络可以识别时间知识维度中的关系,而无须像 CNN 网络中那样组合时间步长^[15]。

CNN 网络与 LSTM 网络相结合的方法主要是利用 CNN 网络在特征提取方面的鲁棒性,以及 LSTM 模型对序列数据中时间依赖性的有效建模,实现了更好的分类性能^[16-17]。Hammerla 等^[18]也尝试将 CNN 网络与 GRU(Gated Recurrent Unit)网络相结合,为了在训练期间捕获不同的数据,将数据输入 3 个独立的通道中,分别使用不同大小的卷积核对数据进行 1DCNN 提取特征,经过最大池化后将数据输入 GRU 网络中捕捉数据的时间依赖性,最后将 3 个独立通道的数据拼接后进行分类。Nafea 等^[19]提出使用 CNN 和 BiLSTM(Bi-Directional Long Short-Term Memory, BiLSTM)并行地提取数据特征,将行为数据分别输入两个独立的网络中,最终将两个子网络的特征进行拼接,完成行为识别。

Khan 等^[20]提出了一种基于注意力的行为识别模型,该框架包含 3 个轻量级的卷积核,从原始数据中提取特征,同时引入了轻量级的多头注意力(Multi-Head Attention)模型,实现了对显著特征的自动选择和对不重要特征的忽略,获得了较好的识别精度。Muhammad 等^[21]提出了一种基于 BiLSTM 的注意力机制分类算法,该机制通过 CNN 提取有效特征,并加入残差模块以保留更多信息的特征,再将这些特征输入到 BiLSTM 中,以学习长期依赖关系,最后结合注意力机制完成行为分类。

元学习^[22-23]的目标是通过一部分任务学习并归纳出一些通用的知识,并利用这些知识在相似但不相同的新任务上仅使用少量样本即可高效地达到较好的效果。具体而言,元学习首先在若干个任务上进行学习,这些任务的性质相似,如在多个图片分类任务上但每个任务所包含的图片种类完全不一样,这样学习出来的模型并不适用于任何一个独立的任务,但具备较高的泛化能力,即较为通用;使用这个较为通用的模型,在一个新的图片分类任务上进行训练,可以快速达到较好的分类效果,一般而言这个新任务具备数据量较小且不易扩大真实数据量的特点。

为了使模型具备较高的泛化能力,元学习方法需要从多个元训练任务中进行学习,获取先验知识以便能够快速适应新任务。在基于优化的元学习方法中,Ravi 等提出了 Meta-LSTM^[24],并发现一般的梯度下降算法的更新过程与 LSTM 算法的更新过程具有较高的相似度,提出将两者结合,引入 LSTM 单元更新公式对网络参数进行更新,在更新过程中将 LSTM 作为一个特殊的记忆器,捕捉参数多次更新的规律,以此为一个新任务提供指导,达到快速更新的目的。Finn 等^[11]

提出了一种基于优化的元学习方法,即 MAML 模型, MAML 的最终目的是通过训练阶段的学习归纳总结出一个泛化能力高的初始化参数。因此,在面对新任务时,模型可以直接使用这个较好的初始化参数,而不必重新训练模型。对于超参数优化的元学习方法可理解为双层优化问题,在训练过程中通过优化多个训练任务来获得多个特定模型参数,并对初始参数进行更新;在测试阶段使用更新过的初始参数在新任务上进行优化更新,以达到快速更新的目的。

MAML 因具备模型无关的特点,可以结合多种网络模型处理不同类型的问题,如 Yao 等^[25]提出了 Meta-ST 网络,使用 ST-Net 网络作为特征提取器,通过 MAML 的更新思想训练得到一组较好的初始化参数,并在新的任务上取得了较高的精度。Feng 等^[26]提出了一种基于度量的小样本学习方法,使用一层 LSTM 进行特征提取,在元训练阶段和元测试阶段使用不同类别的行为动作进行实验,以达到快速学习新的行为类别的目的。Wijekoon 等^[9]提出了基于 MAML 的

行为识别算法,并对比了 RN(Relation network, RN)^[27]、MN 等多种方法;Wijekoon 还提出了以“用户-动作”作为标签的设置,在行为识别问题上进行了新的定义。其实验结果证明在元学习设置下的行为识别精度高于传统训练方法,但在实验中其仅使用了全连接层作为特征提取器,过于简单的网络容易受到训练数据的影响,这无法完全体现出元学习在用户行为识别上的优越性。MAML-M 实验采用了更复杂的模型,但更复杂的模型与 MAML 方法的结合却在小样本的情况下取得了更好的效果,这更能证明元学习的适用性、有效性以及优越性。

3 MAML-M 模型

本文提出的 MAML-M 模型的结构如图 1 所示。该模型主要由 3 个部分构成,依次为用于提取数据特征的 CovLSTM 模块、对特征进行筛选的 Memory 模块以及用于知识归纳的 MAML 方法。

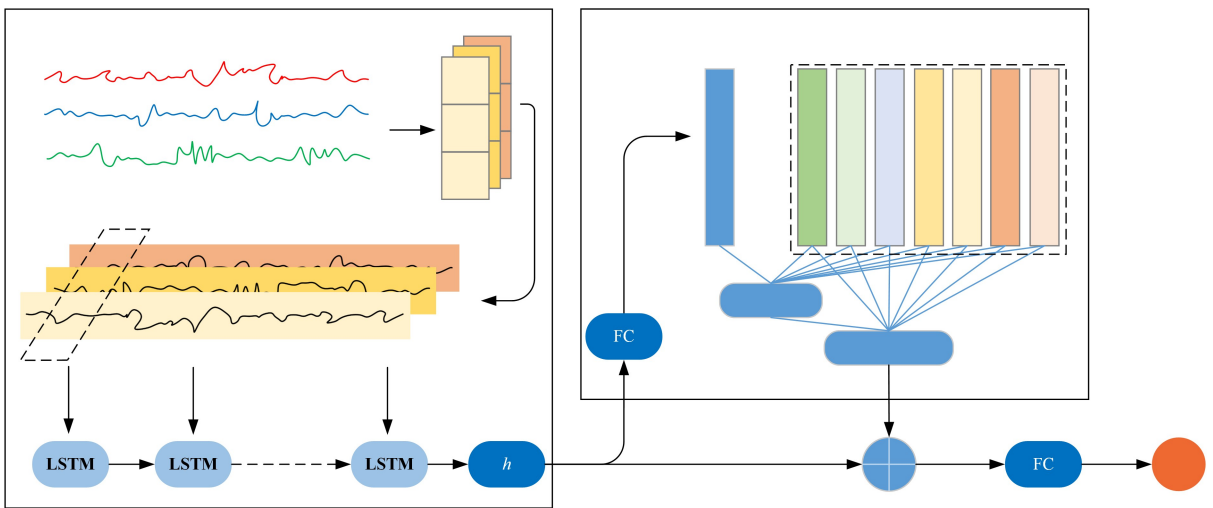


图 1 MAML-M 模型结构图

Fig. 1 MAML-M model structure diagram

3.1 CovLSTM 模块

对于传感器数据,首先使用 CovLSTM 网络对其进行分析和表征,卷积神经网络能够有效地捕捉原始数据在空间上的依赖性,而 LSTM 网络则能有效地提取序列数据在时间维度上的依赖性,通过这样的组合模型,可以更好地将数据的时空特征进行筛选和提取,进而增强模型的分类效果。首先将序列数据输入 CNN 网络中进行特征提取,卷积公式如式(1)所示:

$$x'_{i,j} = \sum_{m=0}^2 \sum_{n=0}^{N-1} W_{m,n} x_{i+m,j+n} \quad (1)$$

卷积核大小设置为 $3 \times N$,其中 3 代表原始数据来自加速度传感器的 3 个轴,即 3 个通道;N 代表卷积核的长度,在本文实验中 $N=1$,同时设置卷积核数量为 3,卷积步幅 $S=1$,不进行 0 填充。此时将卷积过后的数据输入 LSTM 网络中,获得隐藏状态 h_t ,完成初步的特征提取。FC(Full Connected Layer)为全连接层网络。

3.2 Memory 网络模块

注意力机制方法通过巧妙地计算流程,提取全局数据中的部分高价值数据,而相关程度不高的信息则被抑制甚至丢弃。具体来说,注意力机制的本质即是对相关内容

进行提取和融合。本文将引入一个独立的矩阵,以外部记忆器的方式对行为数据的类别特征进行记忆,注意力机制则是将基础网络部分的输出与记忆器作为观察域进行特征分析,这个外部记忆器即是 Memory 模块,如图 2 所示。

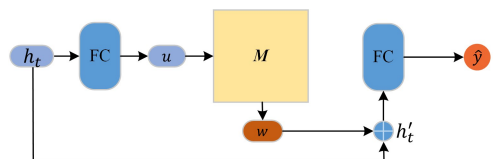


图 2 Memory 模块的处理流程

Fig. 2 Memory module processing flow

从模块的流程图可以看出,当 LSTM 网络完成对数据的分析和表征后,输出最后一次计算的隐藏状态 h_t ,此时先将 h_t 输入到一层全连接网络中,得到输出结果 u ,表达式如式(2)所示:

$$u = W_{fc} h_t + b_{fc} \quad (2)$$

对于 Memory 模块,矩阵维度设置为 $D \times E$,即 $M \in \mathbb{R}^{D \times E}$,其中 D 为分类器的类别数量; E 为特征维度,可以作为一个超参数。Memory 模块的具体结构如图 3 所示。

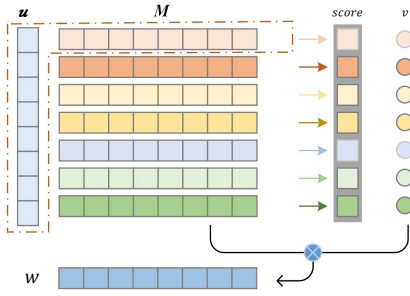


图3 Memory结构

Fig. 3 Memory structure

经过全连接层的表征向量 u 和记忆器 Memory 的每一行向量进行注意力得分计算,这里使用点乘法,即向量内积来计算得分,表达式如式(3)所示:

$$\text{score}(u, M(d)) = \text{dot}(u, M(d)) \quad (3)$$

求得所有的注意力得分后,对得分进行一次归一化计算,即可得出记忆器中需要被重点关注的数据,计算式如式(4)所示:

$$v(d) = \frac{\exp(\text{score}(u, M(d)))}{\sum_{d'} \exp(\text{score}(u, M(d)))} \quad (4)$$

将记忆器的行向量的得分与相应的行向量进行计算,再将所得出的所有加权过后的行向量进行求和,即可得到新的表征向量 w ,表达式如式(5)所示:

$$w = \sum_d v(d) * M(d) \quad (5)$$

最后将得到的表征向量 w 与隐藏状态 h_t 进行拼接,得到新的表征向量 h_t' ,表达式如式(6)所示:

$$h_t' = [h_t; w] \quad (6)$$

最后的分类过程如式(7)所示:

$$\hat{y} = \text{Softmax}(W_{fc} h_t' + b_{fc}) \quad (7)$$

需要特殊说明的是,为了更好地训练新的基础网络部分,即 CNN 网络、LSTM 网络以及 Memory 网络,本文使用了额外的损失函数,这个损失函数可以对 Memory 模块起到更好的训练效果,我们将在下文进行详细说明。

矩阵 M 的维度设定为 $D \times E$,其中 D 设为分类器的类别数量,即我们期望 M 中的每一行能代表一种行为动作的特征,这样记忆器 M 在与隐藏状态 h_t 进行注意力计算后,就可以根据注意力打分来初步确定该样本的所属类别。因为注意力得分是通过向量内积的方法进行计算的,结果越大,代表其特征相似程度越高,即越有可能是该行向量所代表的动作类别。因此,需要将隐藏状态 h_t 所代表样本的真实类别与记忆器 M 中行向量所代表的动作类别联系起来,才能确保记忆器 M 起到我们预期的效果。

首先,需要在训练过程中加入新的损失函数,如式(8)所示:

$$\mathcal{L}_{p_i^{\text{train}}}(\mathbf{M}) = - \sum_{(x, y) \in p_i^{\text{train}}} O(y) \log v \quad (8)$$

其中,向量 v 代表隐藏状态与记忆器各行向量的注意力得分向量,此时的得分已经进行过归一化计算; y 表示输入样本 x 的真实标签, $O(y)$ 则代表真实标签转换后的 one-hot 向量。

由新的损失函数可以看出,可以通过真实标签 one-hot 向量锁定记忆器中该类别所对应的行向量,并将此损失加入

到目标函数中进行优化,从而达到我们的预期目标。

3.3 MAML 方法

MAML 是基于优化的元学习中最经典的方法,其核心思想是通过跨任务训练策略,在多个元训练任务上学习归纳出一个良好的初始化参数,使得模型能够以此为基础快速适应新任务。总的来说,MAML 方法的目标是获得一个良好的初始化参数 θ ,这个初始化参数并不需要在某一特定任务上表现良好,而是要求适用于各种相似的任务,即具备较强的泛化能力,整体网络框架如图 4 所示。

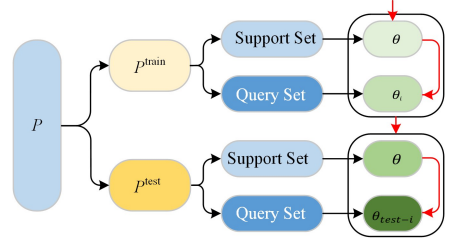


图4 MAML网络框架

Fig. 4 MAML network framework

根据元学习的设定,元训练过程需要多个训练任务,因此首先定义用户集 \mathcal{P} ,用户集中包含多个用户 p , $\mathcal{P} = \{p_i\}$ 即 $\mathcal{P} = \{p_1, p_2, \dots, p_i, \dots, p_{N_{All}}, 1 \leq i \leq N_{All}\}$,每个用户 p 包含其所有的行为数据, N_{All} 代表用户数量;将用户集 P 随机划分为 $\mathcal{P}^{\text{train}}$ 和 $\mathcal{P}^{\text{test}}$,将其分别作为用于元训练和元测试的用户集合;用户行为数据的类别数量为 N ,定义类别集合 $C, C = \{c_1, c_2, \dots, c_N\}$ 。

在用于元训练的用户集 $\mathcal{P}^{\text{train}}$ 中,包含多个用户 p_i^{train} ,即 $\mathcal{P}^{\text{train}} = \{p_1^{\text{train}}, p_2^{\text{train}}, \dots, p_i^{\text{train}}, \dots, p_{N_s}^{\text{train}}, 1 \leq i \leq N_s\}$, N_s 为用于元训练的用户集的用户个数,每个用户包含一组行为数据,即包含 N_s 组数据,每组数据独属于一名用户。

用户数据 p_i^{train} 中包含该用户的全部行为数据,即 $p_i^{\text{train}} = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_{N_i}, y_{N_i})\}$,其中 N_i 代表用户任务 p_i^{train} 中数据的数量, $y_i \in C$ 。根据元学习的设定,将 p_i^{train} 中的用户行为数据划分为支持集 \mathcal{D}^s 和查询集 \mathcal{D}^q 。在本次实验中,根据类别集合 C ,针对用户数据 p_i^{train} 中的数据,按行为类别从每类中随机抽取 K^s 条数据,构成支持集 p_{i-s}^{train} ,可表示为 $p_{i-s}^{\text{train}} = \{(x_{c_1,1}^s, c_1), (x_{c_1,2}^s, c_1), \dots, (x_{c_1,K^s}^s, c_1), (x_{c_2,1}^s, c_2), \dots, (x_{c_2,K^s}^s, c_2), \dots, (x_{c_N,K^s}^s, c_N)\}$,即一共 N 个类别,共包含 $N \times K^s$ 条数据;在剩余的数据中,每个类别随机抽取 K^q 条数据,构成查询集 p_{i-q}^{train} ,可表示为 $p_{i-q}^{\text{train}} = \{(x_1^q, y_1), (x_2^q, y_2), \dots, (x_{K^q}^q, y_{K^q})\}$ 。在此定义用户任务 $p_{i-task}^{\text{train}} = p_{i-s}^{\text{train}} + p_{i-q}^{\text{train}}$, $p_{i-s}^{\text{train}} \cap p_{i-q}^{\text{train}} = \emptyset$,用户任务的数据由该用户的全部数据 p_i^{train} 随机抽取组成,用户任务中的支持集 p_{i-s}^{train} 和查询集 p_{i-q}^{train} 数据完全不同。上述过程完成了一项用户任务的构建,包含相应的支持集和查询集。对比传统元学习设定中的“ N -way- K -shot”,本实验使用 N 种行为类别,每种类别使用 K^s 条数据构成支持集, K^q 条数据构成查询集,每项用户任务共包含 $N \times (K^s + K^q)$ 条数据,在实验中, $K^s = 5, K^q = 15$ 。

重复以上方法,用于元训练的用户集 $\mathcal{P}^{\text{train}}$ 最终构建出元训练集 p^{train} ,采用同样的方式,对于用于元测试的用户集 $\mathcal{P}^{\text{test}}$ 最终构建出元测试集 p^{test} 。算法 1 对具体流程进行了详细描述。

算法 1 元训练任务生成过程

输入:用户数据集 $\mathcal{P} = \{p_1, p_2, \dots, p_i, \dots, p_{N_{All}}, 1 \leq i \leq N_{All}\}$, 类别集合

C , 参数 K^s 和 K^q , 元训练任务数量 N_s .

输出:元训练集 p_{task}^{train}

1. $\mathcal{P}^{train} = \text{randomSample}(\mathcal{P}, N_s)$
2. For $p_i^{train} \sim \mathcal{P}^{train}$
3. $p_{i-s}^{train} = \text{randomSample}(p_i^{train}, C, K^s)$
4. $p_{i-q}^{train} = \text{randomSample}((p_i^{train} - p_{i-s}^{train}), C, K^q)$
5. $p_{i-task}^{train} = p_{i-s}^{train} + p_{i-q}^{train} / * p_{i-s}^{train} \cap p_{i-q}^{train} = \emptyset /$
6. $p_{task}^{train} \leftarrow p_{i-task}^{train}$
7. End

在元训练阶段,先使用元训练任务的支持集数据对初始模型进行训练更新,此时使用的是随机初始化的参数,损失函数的计算式如式(9)所示:

$$\mathcal{L}_{p_i^{train}}(f_{\theta_i'}) = \sum_{(x,y) \in p_i^{train}} y \log f_{\theta_i'}(x) \quad (9)$$

通过求得的损失,对模型参数进行梯度下降更新,计算式如式(10)所示。此时更新的参数不但包括卷积神经网络和循环神经网络中的所有参数,还包括记忆器部分的参数,即对所有的参数 θ' 进行更新,任务学习率为 α 。

$$\theta_i' = \theta' - \alpha \nabla_{\theta'} \mathcal{L}_{p_i^{train}}(f_{\theta'}) \quad (10)$$

新得到的参数 θ_i' 则是通过该用户任务的支持集 p_{i-s}^{train} 训练得到的适合用户的参数,继续在此模型上用该用户任务的查询集进行训练,即可得到新的损失 $\mathcal{L}_{p_i^{train}}(f_{\theta_i'})$, 计算式如式(11)所示:

$$\mathcal{L}_{p_i^{train}}(f_{\theta_i'}) = \sum_{(x,y) \in p_i^{train}} y \log f_{\theta_i'}(x) \quad (11)$$

此时,在计算 $\mathcal{L}_{p_i^{train}}(f_{\theta_i'})$ 的同时,还要对新的损失 $\mathcal{L}_{p_i^{train}}(M)$ 进行计算,计算式如式(12)所示。 $\mathcal{L}_{p_i^{train}}(M)$ 即是在该用户任务的查询集上的损失,用于加强对记忆器模块的更新,使其能够更好地对特征向量进行再次表征。

$$\mathcal{L}_{p_i^{train}}(M) = - \sum_{(x,y) \in p_i^{train}} O(y) \log v \quad (12)$$

为了提高计算效率,加快收敛速度,我们利用多个用户任务在查询集上的损失和来对参数 θ' 进行更新,表达式如式(13)所示:

$$\theta' \leftarrow \theta' - \beta \nabla_{\theta'} \sum_{p_i^{train}} \mathcal{L}_{p_i^{train}}(f_{\theta'}) - \gamma \sum_{p_i^{train}} \mathcal{L}_{p_i^{train}}(M) \quad (13)$$

其中, γ 记为记忆学习率,通过新的计算公式对 θ' 进行更新,此时得到的参数 θ' 记为元学习的目标,一个能够在新用户的行为识别任务上快速适应的良好参数,此时完成了元训练阶段的计算。

在元测试阶段,使用元测试集中用户任务的支持集对模型进行更新,即在元学习模型学习到的良好的初始化参数 θ' 上进行再次训练,此时面对新的用户任务,模型计算出的损失如式(14)所示:

$$\mathcal{L}_{p_i^{test}}(f_{\theta'}) = \sum_{(x,y) \in p_i^{train}} (y \log f_{\theta'}(x) + (1-y) \log(1 - f_{\theta'}(x))) \quad (14)$$

通过求得的损失 $\mathcal{L}_{p_i^{test}}(f_{\theta'})$, 对模型参数进行梯度下降更新,表达式如式(15)所示:

$$\theta'_{test-i} = \theta' - \alpha \nabla_{\theta'} \mathcal{L}_{p_i^{test}}(f_{\theta'}) \quad (15)$$

新得到的参数 θ'_{test-i} 则是通过该用户任务的支持集 p_{i-s}^{test} 训练得到的最适合该用户的参数,且是在由元训练阶段所得出的良好的初始化参数 θ' 的基础上训练得来的。此时使用

θ'_{test-i} 对该任务的查询集进行测试,即得到这些样本的预测分类。元学习框架的整个训练和测试流程到此结束, θ'_{test-i} 即是新用户的行为识别任务的最优模型参数。MAML-M 元学习过程如算法 2 所示。

算法 2 MAML-M 元学习过程

输入:元训练任务 p_{task}^{train} , 元测试任务 p_{task}^{test} , 任务学习率 α , 元学习率 β , 记

忆学习率 γ

输出:元测试任务 p_{task}^{test} 中查询集行为识别结果

1. 随机初始化参数 θ' , $\theta' = \{\theta \cup \theta_M\}$ 。
2. while not done
3. /* 元训练过程 */
4. sample a batch of tasks from p_{task}^{train}
5. for each task do:
6. 使用该训练任务的支持集数据计算损失 $\mathcal{L}_{task}(f_{\theta'})$, 如式(9)所示。
7. 使用该损失更新参数 θ' , 如式(10)所示。
8. 使用该训练任务的查询集数据计算损失 $\mathcal{L}_{task}(f_{\theta'})$, 如式(11)所示。
9. 使用该训练任务的查询集数据计算损失 $\mathcal{L}_{task}(M)$, 如式(12)所示。
10. end
11. 使用该批次任务的查询集损失和更新参数 θ' , 如式(13)所示。
12. end
13. for each task from p_{task}^{test}
14. /* 元测试过程 */
15. 使用该测试任务的支持集数据计算损失 $\mathcal{L}_{task}(f_{\theta'})$, 如式(14)所示。
16. 使用该损失更新参数 θ' , 如式(15)所示。
17. 使用更新后的参数 θ'_{test-i} 对测试任务的查询集数据进行预测。
18. end

4 实验**4.1 数据集**

MEx^[28] 是一个理疗运动数据集,包含的几种动作经常被用于预防或治疗肌肉骨骼疼痛。文献[28]对每个动作的起始位置和动作规范都做了说明。该数据集由 30 名志愿者提供,每位志愿者分别进行 7 种不同的理疗运动,如转膝、俯卧伸展、挥拳等动作,具体的“编号-动作”包括:01-Knee-rolling, 02-Bridging, 03-Pelvic tilt, 04-The Clam, 05-Repeated Extension in Lying, 06-Prone punches, 07-Superman。其中,用户在手腕(ACW)和大腿上(ACT)分别穿戴了加速度计,并根据其提供的传感器数据流创建了两个不同的数据集。参与者的信息如图 5 所示,图 5 中统计了数据提供者的年龄和相应人数的分布。

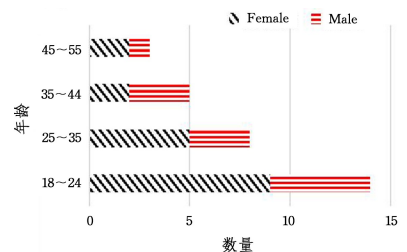


图 5 参与者信息统计

Fig. 5 Participant information statistics

下文对实验中的几种对比方法进行介绍。

ANN^[28]:简单地感知神经网络,由一层 100 个神经元的隐藏层构成特征提取网络,通过批归一化后传入 Softmax 层进行分类。

CNN^[28]:卷积神经网络,将输入数据进行两次一维卷积和最大池化操作,并通过两层全连接层计算,最后传入 Softmax 层进行分类。

CovLSTM^[28]:将输入数据进行一次一维卷积后进行维度置换,继续将数据输入 LSTM 网络,并通过两层全连接层进行计算,最后传入 Softmax 层进行分类。

MN^[9]:匹配网络,通过一层 1 200 个神经元的隐藏层进行特征提取并进行批归一化,计算两个样本的余弦相似度并进行分类。

MAML-DNN^[9]:使用 MAML 框架的感知神经网络,由一层 1 200 个神经元的隐藏层构成。

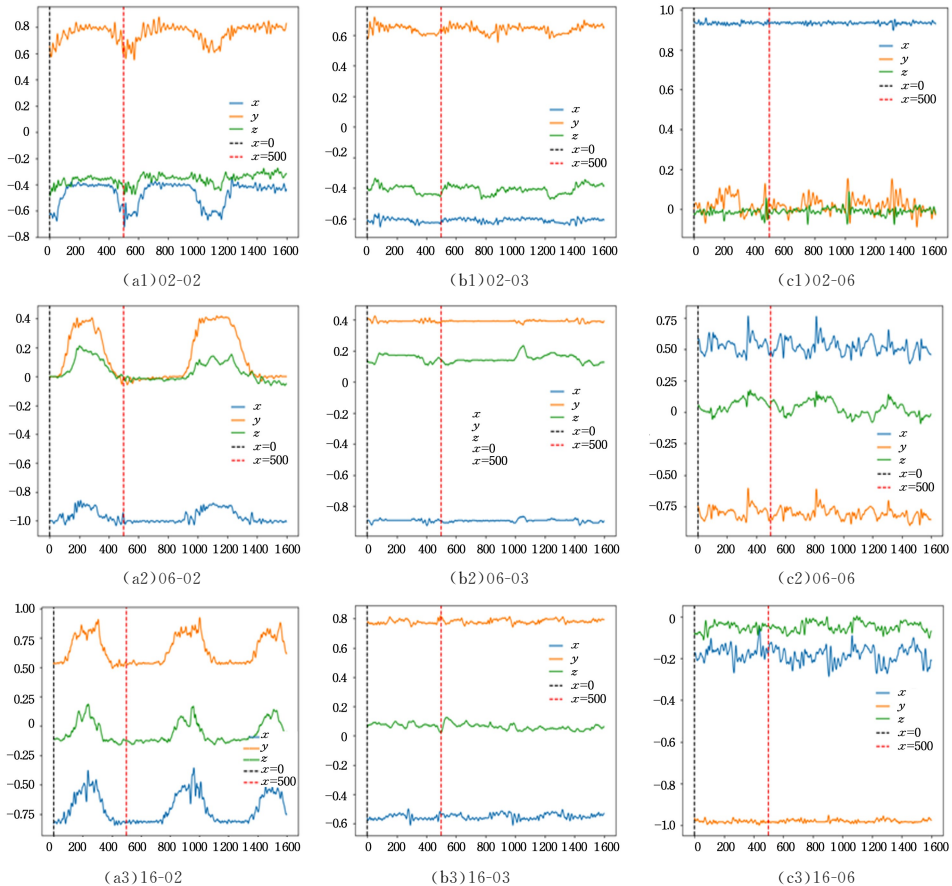
MAML-CovLSTM:未添加 Memory 模块的 MAML-M 网络,用于进行对比。

MAML-M-noLossM:加入了 Memory 模块,但未对 Memory 模块添加新的损失,与 MAML-M 网络进行对比。

4.2 实验设置

对于 MAML-M 模型,本文设置任务学习率 α 为 10^{-2} ,元学习率 β 为 10^{-3} ,记忆学习率 γ 为 10^{-4} ;每次元训练的任务数设置为 4,以加快模型的收敛。

在 MEx 数据集中,每个用户的每个动作类别所提取的



注:MEx-ACT 行为数据片段以“志愿者编号-动作编号表示”。

图 6 MEx-ACT 行为数据片段

Fig. 6 MEx-ACT behavior data fragment

行为动作样本片段数量相近,即在构成查询集的样本中类别较为平均,因此可以使用分类正确率作为模型的评价指标,如式(16)所示:

$$Acc = \frac{\sum_i^K f(x_i) = y_i}{K^q} \quad (16)$$

其中, K^q 表示测试集中每个类别的样本数, $f(x_i)$ 表示预测类别, y_i 为样本的真实类别。

4.3 行为数据分析

随机抽取部分参与者的部分行为类别的数据进行可视化分析,行为数据抽取情况如表 1 所列。在各个数据集中,每位志愿者提供的行为数据以“编号-动作”表示。

表 1 数据抽样

Table 1 Data sampling

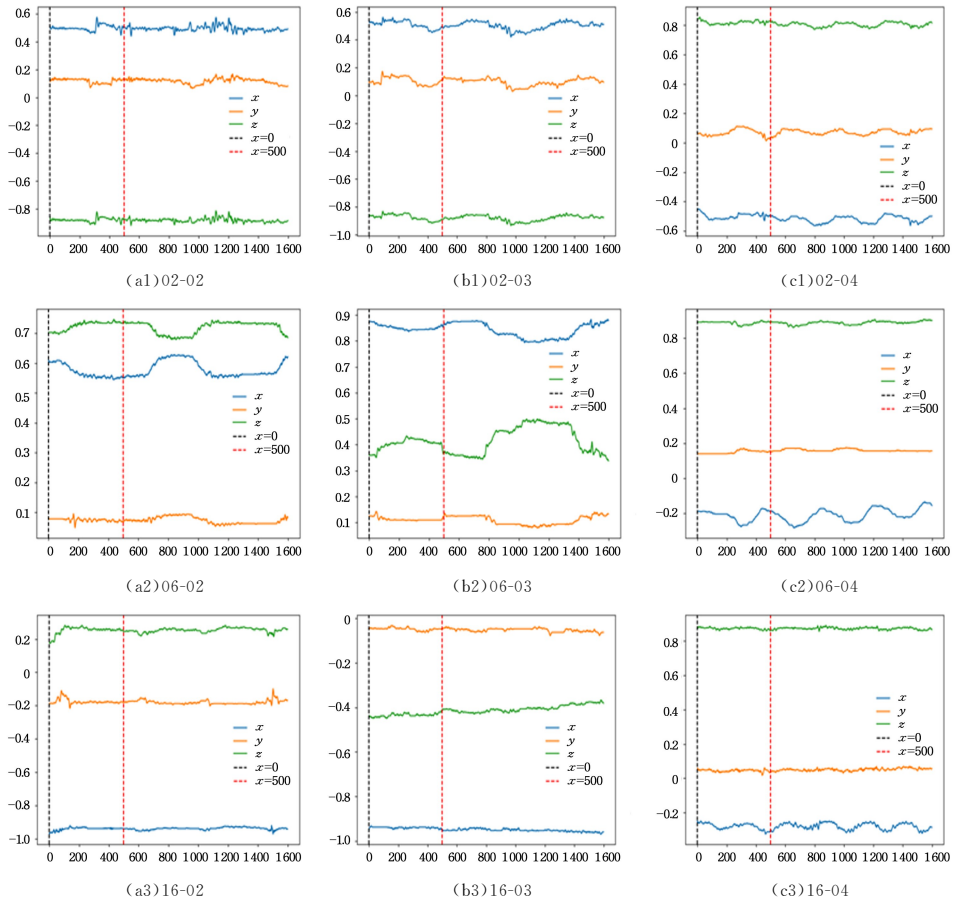
数据集	人员编号		
	02	06	16
MEx-ACT	02-Bridging;03-Pelvic tilt;06-Prone punches		
Mex-ACW	02-Bridging;03-Pelvic tilt;04-The Clam		

为了更加清晰地检测原始输入数据的数据可用性,对部分数据片段进行了可视化分析,展示数据片段长度为 1 600,并根据滑动窗口大小进行了辅助线标注,直观显示提取的数据片段。其中,在 MEx-ACT 数据集中的行为数据可视化如图 6 所示,其中纵轴为传感器加速度值(m/s^2),横轴为时间序列采样点(个),采样频率为 100 Hz。

由图 6 可看出,同一用户的不同行为有着较明显的差别,具备训练模型、实现有效分类的数据基础。但不同用户的同一动作也存在着明显的差异,即同类别的行为数据呈现出较大差异。根据可视化分析,这种差异除了来自个体的行为习惯外,还有可能来自传感器采集设备的佩戴方式,如在 06-

Prone punches 动作中,明显存在不同数据轴的数据在分布位置上的交换,这一点在 MEX-ACW 数据集中表现得更为明显。

在 MEX-ACW 数据集中的行为数据可视化如图 7 所示,在数据集中提取了编号为 02,06,16 的 3 位志愿者的数据,行为类别为 02-Bridging,03-Pelvic tilt 以及 04-The Clam。



注: MEX-ACW 行为数据片段以“志愿者编号-动作编号表示”。

图 7 MEX-ACW 行为数据片段

Fig. 7 MEX-ACW behavior data fragment

由图 7 可看出,有些用户的不同行为有着较为明显的差别,但也有用户在进行不同行为时产生了极为相似的数据,如 06 号志愿者的 02-Bridging 和 03-Pelvic tilt 两个动作,所产生的数据相似度较高,这为模型的行为识别分类带来了较大的挑战。该数据集的传感器数据提取自绑定在大腿的传感器,分析其具体动作,这两种行为的下肢动作幅度较小,对比其他志愿者,仍存在类似情况。不同用户的同一动作也存在着明显的,如在 02-Bridging 动作中,02 号志愿者和 06 号志愿者的动作数据存在着明显差异,两位志愿者的 X 轴和 Z 轴的行为数据分布发生了对换,这可能是佩戴方式不同所造成的,如将传感器绑在大腿正面或绑在大腿侧面,由于该动作为平躺类型,所呈现出的数据差异可以得到合理解释,排除噪声样本的可能。

类别的行为数据,每种动作类别的支持集有 5 条样本,查询集有 15 条样本,每个用户任务包含 140 条数据。在 MEX-ACT 数据集中,实验构建了 100 个 epoch,测试过程的正确率曲线如图 8 所示。

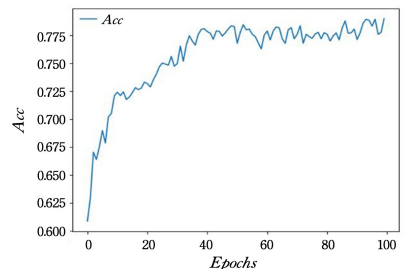


图 8 元测试过程正确率曲线

Fig. 8 Meta test process correctness curve

4.4 实验及结果分析

本次实验通过元学习设定的标准进行数据采样的任务划分,即针对数据集构建 7-way 5-shot 分类任务。在用户任务中,设置查询集样本数量为 15,即每个用户任务包含 7 个

在测试过程中,将每轮训练得到的参数用测试集进行测试,取首次测试的结果,即首次在测试集数据上进行分类结果准确率的预测。该准确率能够反映出元学习方法通过大量训练任务所学习到的初始化参数的泛化性能,一个好的初始化

参数才能使模型快速适应新任务。可以看到,模型在面对新任务时,初始参数能够达到较好的分类效果,在新任务上达到了较好的分类表现。由图 8 中的数据曲线可以看出,在训练第 100 个 epoch 后初始化参数的性能增长趋于稳定,最优达到 78.9%。

在测试过程中,为了更好地评估初始化参数在新任务上的适应能力,每次测试进行 10 步梯度下降更新,即在测试集上训练 10 轮。训练第 100 轮后的初始化参数测试过程的正确率曲线如图 9 所示。可以看到,模型在面对新任务时,在不更新参数的情况下就能够达到较好的分类效果,随着几次梯度下降更新后,模型在新任务上达到了较好的分类表现,准确率达到了 91.27%,证明了模型在新任务上的快速适应能力。

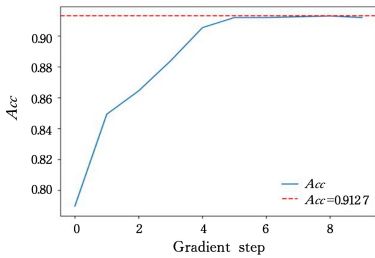


图 9 元测试过程更新正确率曲线

Fig. 9 Meta test process update correctness curve

表 2 列出了在 MEX-ACT 和 MEX-ACW 数据集上进行小样本行为识别分类的测试表现,实验进行了 7-way-5-shot 的实验设置,本文提出的 MAML-M 与 MAML-CovLSTM 等方法的实验结果对比如表 2 所列。

表 2 MEX-ACT 数据集的实验结果(Acc)

Table 2 Experimental results of MEX-ACT dataset(Acc)

方法	数据集	
	MEX-ACT	MEX-ACW
ANN	0.8726	0.5741
CNN	0.8704	0.5754
CovLSTM	0.8892	0.6499
MN	0.9073	0.4620
MAML-DNN	0.8673	0.6525
MAML-CovLSTM	0.9095	0.6727
MAML-M-noLossM	0.8906	0.6629
MAML-M	0.9127	0.6908

结果表明,本文提出的 MAML-M 模型在两个数据集上取得了较高准确率,并且超过了 MAML-CovLSTM 模型和其他对比方法,这证明了元学习框架在面对新任务时不但有较好的泛化能力,还能够快速达到预期的训练效果;同时 Memory 模块也有效地提升了模型的分类效果。通过比较 MAML-M-noLossM 和 MAML-M 可以发现,增加了 Memory 模块的损失后,通过训练可以得到更好的 Memory 模块参数,能够有效地提升模型效果。通过对比实验数据可以发现:对比 MAML-CovLSTM 方法,MAML-M 模型取得了更好的效果,在 MEX-ACT 和 MEX-ACW 两个数据集上的准确率分别提高了 0.32% 和 1.81%,这代表模型具备更强的特征提取能力和分类能力,为 MAML 框架提供了更好的学习基础。两个实验的测试结果如图 10 所示。

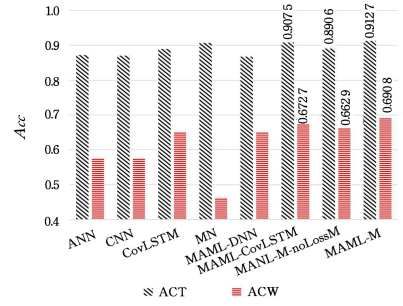


图 10 测试结果对比

Fig. 10 Comparison of test results

结束语 本文提出了一种基于元学习和注意力机制的用户行为识别模型 MAML-M。该模型一方面利用 CovLSTM 特征提取器和注意力机制来有效提高模型的分类效果,另一方面利用元学习的方法学习通用模型初始化参数,使模型仅使用新用户的少量数据便能快速更新到较高的精度,这有效地解决了用户个性化差异对动作识别产生的问题。MAML-M 模型首先利用 CovLSTM 模块提取动作特征,然后通过 Memory 模块提取重要特征,达到提升分类效果的目的,最后使用 MAML 元学习方法来学习先验知识并进行习惯归纳总结,得到泛化能力较高的初始化参数,提高了新用户的动作识别准确率。未来的研究方向主要为:1)考虑对比不同的小样本学习方法,寻找最优的解决方案;2)研究新用户新行为动作类别的小样本问题。

参考文献

- [1] SABOOR A, KASK T, KUUSIK A, et al. Latest research trends in gait analysis using wearable sensors and machine learning: A systematic review[J]. IEEE Access, 2020, 8: 167830-167864.
- [2] SHIH C S, CHOU J, LIN K J. WuKong: Secure Run-Time environment and data-driven IoT applications for Smart Cities and Smart Buildings[J]. Journal of Internet Services and Information Security, 2018, 8(2): 1-17.
- [3] ZHANG S, LI Y, ZHANG S, et al. Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances[J]. Sensors, 2022, 22(4): 1476.
- [4] 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.
- [5] MEKRUKSANICH S, JITPATTANAKUL A. LSTM networks using smartphone data for sensor-based human activity recognition in smart homes[J]. Sensors, 2021, 21(5): 1636.
- [6] JOBANPUTRA C, BAVISHI J, DOSHI N. Human activity recognition: A survey[J]. Procedia Computer Science, 2019, 155: 698-703.
- [7] DARGAN S, KUMAR M, AYYAGARI M R, et al. A survey of deep learning and its applications: a new paradigm to machine learning[J]. Archives of Computational Methods in Engineering, 2020, 27(4): 1071-1092.
- [8] WANG Y, YAO Q, KWOK J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing

- Surveys(csar),2020,53(3):1-34.
- [9] WIJEKON A,WIRATUNGA N. Personalised meta-learning for human activity recognition with few-data[C]// International Conference on Innovative Techniques and Applications of Artificial Intelligence. Cham:Springer,2020:79-93.
- [10] HUISMAN M,VAN RIJN J N,PLAAT A. A survey of deep meta-learning[J]. Artificial Intelligence Review,2021,54(6):4483-4541.
- [11] FINN C,ABBEEL P,LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]// International Conference on Machine Learning. PMLR,2017:1126-1135.
- [12] LECUN Y,BOTTOU L,BENGIO Y,et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE,1998,86(11):2278-2324.
- [13] 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.
- [14] BALDOMINOS A,CERVANTES A,SAEZ Y,et al. A comparison of machine learning and deep learning techniques for activity recognition using mobile devices[J]. Sensors,2019,19(3):521.
- [15] ZHAO Y,YANG R,CHEVALIER G,et al. Deep residual bidirectional LSTM for human activity recognition using wearable sensors [J]. Mathematical Problems in Engineering,2018,2018:7316954.
- [16] MUTEGEKI R,HAN D S. A CNN-LSTM approach to human activity recognition[C]//2020 International Conference on Artificial Intelligence in Information and Communication(ICAIIIC). IEEE,2020:362-366.
- [17] ORDÓÑEZ F J,ROGGEN D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition[J]. Sensors,2016,16(1):115.
- [18] HAMMERLA N Y,HALLORAN S,PLÖTZ T. Deep, convolutional, and recurrent models for human activity recognition using wearables[J]. arXiv:1604.08880,2016.
- [19] NAFEA O,ABDUL W,MUHAMMAD G,et al. Sensor-based human activity recognition with spatio-temporal deep learning [J]. Sensors,2021,21(6):2141.
- [20] KHAN Z N,AHMAD J. Attention induced multi-head convolutional neural network for human activity recognition[J]. Applied Soft Computing,2021,110:107671.
- [21] MUHAMMAD K,ULLAH A,IMRAN A S,et al. Human action recognition using attention based LSTM network with dilated CNN features[J]. Future Generation Computer Systems,2021,125:820-830.
- [22] THRUN S,LORIEN P,et al. Learning to learn[M]. Springer Science & Business Media,2012.
- [23] LAKE B M,SALAKHUTDINOV R,TENENBAUM J B. Human-level concept learning through probabilistic program induction[J]. Science,2015,350(6266):1332-1338.
- [24] RAVI S,LAROCHELLE H. Optimization as a model for few-shot learning [C]// International Conference on Learning Representations,2017.
- [25] YAO H,LIU Y,WEI Y,et al. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction[C]// The World Wide Web Conference,2019:2181-2191.
- [26] FENG S,DUARTE M F. Few-shot learning-based human activity recognition[J]. Expert Systems with Applications,2019,138:112782.
- [27] SUNG F,YANG Y,ZHANG L,et al. Learning to compare: Relation network for few-shot learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:1199-1208.
- [28] WIJEKON A,NIRMALIE W,KAY C. Mex: Multi-modal exercises dataset for human activity recognition[J]. arXiv:1908.08992,2019.



WANG Jiahao, born in 1978, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include IoT, information security and data mining.



ZHAO Dexin, born in 1984, Ph.D, associate researcher. His main research interests include underwater IoT, intelligent recognition and autonomous underwater vehicles.

(责任编辑:喻黎)