

基于注意力的卷积神经网络模型用于使用 sEMG 的人类下肢活动识别

M. Mollahosseini¹, F.H. Daryakenari², M.H. Rohban³, G.R. Vossoughi^{*1}

Abstract—使用表面肌电图 (sEMG) 信号对下肢运动进行准确分类在辅助机器人和康复系统中起着至关重要的作用。在这项研究中, 我们提出了一种轻量级基于注意力的深度神经网络 (DNN), 用于使用公开的 BASAN 数据集中的多通道 sEMG 数据进行实时运动分类。所提出的模型仅包含 62,876 个参数, 并设计成无需计算量大的预处理步骤, 使其适合于实时部署。我们采用留一法验证策略以确保模型在不同个体间的普适性, 并在三类运动 (行走、屈膝站立和屈膝坐下) 上评估了模型。网络在验证集上实现了 86.74% 精度, 在测试集上实现了 85.38%, 显示出在现实条件下强大的分类性能。与文献中的现有模型进行的比较分析突出了我们方法的效率和效力, 特别是在计算成本和实时响应至关重要的情况下。结果表明, 该模型是集成到人机交互系统上层控制器中的一个有希望的候选者。

Index Terms—Attention Mechanism, sEMG Signals Classification, Deep Neural Network, Lower Limb Rehabilitation.

I. 介绍

表面肌电图 (sEMG) 信号已被广泛应用于各种应用中, 包括人机交互、神经肌肉疾病诊断和康复。其最重要的应用之一是下肢活动识别, 它在医疗监测、主动与辅助系统以及远程沉浸中起着至关重要的作用 [?]. 人类活动识别可以通过视觉传感器或可穿戴传感器来实现 [?]. 可穿戴传感器直接放置在主体的身体上, 而视觉传感器 (如摄像头) 不需要物理附着。然而, 由于视觉传感器的隐私问题, 可穿戴传感器通常更受青睐 [?]. 随着可穿戴传感器技术的进步, 已为人体运动分析提供了各种设备, 如加速度计、肌电电极、陀螺仪和气压计。在这些设备中, 肌电传感器尤其具有优势, 因为它们能直接提供肌肉活动的信息, 从而比其他可穿戴传感器提供更准确的运动识别。

然而, 由于 sEMG 信号的固有特性及其对各种噪声源的敏感性, 这些信号本质上是复杂的。这些信号通常使用放置在皮肤上的表面电极获取, 这导致对潜在肌肉活动的间接测量。由于这种间接性, 使得信号非常容易受到噪声的影响, 尤其是由于肌肉在活动过程中产生的运动伪迹以及肌肉串扰造成的噪声——这是相邻肌肉接近导致的结果。

近年来, 使用 sEMG 信号进行人类活动分类受到越来越多的关注。然而, EMG 信号中的噪声仍然是一个重大的挑战。常见的噪声源包括环境干扰、运动伪影、信号不稳定以及肌群之间的串扰。虽然某些预处理技术可以帮助降低噪声 [4], 但通常需要先前的噪声消除方法来实现高分类精度。几项研究采用了各种技术来减少 sEMG 信号

中的噪声, 包括小波去噪 [?]、独立成分分析 (ICA) 以及经验模态分解 (EMD) [?]. 去噪后, 处理过的信号通常被送入分类器。然而, 一些研究选择使用原始 sEMG 数据 (包括噪声) 以降低计算复杂性和处理时间 [8]。这种方法使得分类器更适合实时应用, 因为它消除了耗时的预处理阶段。为此, 需要一种对噪声具有鲁棒性的分类建模方法。

随着深度学习 (DL) 的最新进展, DL 架构已经成为基于 sEMG 信号设计稳健高级控制器的有效工具 [?]. 正如我们之前的研究 [?] 所示, 卷积神经网络 (CNN) [?] 在这方面特别适合, 因为它们能够自动从原始传感数据中提取有意义的特征。尽管 CNN 传统上用于处理二维数据, 如图像, 但它们在分析一维信号 (包括 sEMG) 方面也被证明是非常有效的。

即使在主要目标是回归的应用中, 例如预测系统的下一个状态, CNN 也可以直接从原始输入信号中提取相关特征, 而不需要手工预处理。例如, 在 [?] 中, 原始 sEMG 信号首先由 CNN 处理, CNN 提取空间特征, 然后将这些特征输入到递归神经网络 (RNN) 中以预测关节角度。这突出了基于 CNN 的架构在人体与机器人交互和控制系统中的分类和回归任务中的多功能性。

虽然深度学习 (DL) 方法展现了很大的潜力, 但实时应用需要在模型精度和计算效率之间进行权衡。在许多情况下, 相较于计算开销大的高精度模型, 更倾向于选择精度较低的模型, 尤其是在实时场景中。为了提高 DL 方法的分类性能, 通常在将 sEMG 信号输入神经网络之前应用降维技术。这些技术有助于提取低维的有效特征, 提高精度和效率。一些常用的方法包括线性判别分析 (LDA)、主成分分析 (PCA)、局部线性嵌入 (LLE)、保秩判别分析和拉普拉斯特征映射 [?].

DL 技术的一个关键优势是其能够自动化特征提取和降维, 消除了手动预处理的需要。虽然有人可能会认为深度学习模型比浅层网络更复杂, 但它们通过直接从原始信号中学习高级表示来提供更好的鲁棒性。值得注意的是, 卷积神经网络 (CNNs) [?]、长短期记忆 (LSTM) 网络 [?]、和深度置信网络 (DBNs) [?] 都被用于基于表面肌电信号 (sEMG) 的动作识别, 并在该领域展示了令人鼓舞的结果 [4]。除了这些著名的 DL 架构之外, 注意机制能够在 DL 方法中创建内和外的数据关系。注意机制 (AM) [?] 是神经网络中的一个附加层, 通过选择性地关注最相关的特征来增强模型性能。当在网络中进行战略性放置时, AM 可以提高训练效率, 并在某些情况下增强可解释性 [?].

虽然在将深度学习建模方法应用于人体活动识别方面取得了许多进展, 但在以往研究中仍存在一个主要的模糊点。这个模糊点是, 在大多数以往的研究中, 他们常常未能明确说明验证和测试数据集是否包含训练阶段中曾出现过的主体。当一个模型在完全未见过的主体上进行评估时, 其性能可能会显著下降。解决该问题的一种方法是留一法, 即在训练集中排除一个主体, 并专门用于测试。这种方法

¹ Faculty of Mechanical Engineering Department, Sharif University of Technology, Tehran, Iran. mojtaba.hoseini1390@yahoo.com, vossough@sharif.edu; ² Faculty of Engineering and IT, the University of Melbourne, Victoria, Australia. f.haghighoodaryakenari@unimelb.edu.au; ³ Computer Engineering Department, Sharif University of Technology, Tehran, Iran rohban@sharif.edu

Digital Object Identifier (DOI): see top of this page.

可以更现实地评估模型对新主体的泛化能力。

在本研究中,我们旨在设计一个基于注意力机制的深度神经网络,其参数数量较少,使其在实时应用中高效,同时在验证集和测试集上达到可接受的准确性。为此,我们将使用一个包含正常受试者和下肢异常受试者的开放源代码数据集来训练我们的模型以识别三类人体运动。在训练、验证和测试过程中,我们将采用留一法策略,以确保报告的结果尽可能具有普遍适用性。最后,我们的方法性能将与类似方法进行比较。

II. 方法

A. 方法概述

在本研究中,我们的目标是提出一种基于注意力的轻量级深度神经网络,用于通过表面肌电图(sEMG)信号分类下肢运动。为此,我们首先利用开源的 BASAN 数据集,该数据集包含从健康个体和膝关节异常患者收集的多通道 sEMG 和膝关节角度数据。对原始信号进行最小的预处理,仅限于重新整形以适应神经网络输入并使用滑动窗口法进行分段。数据采用留一法进行划分,以确保稳健和可推广的评估。所提出的网络结构由多层卷积层和最大池化组成,随后是扁平化和注意力机制,最后达到用于最终分类的密集层。没有进行额外的信号去噪或特征提取,确保与实时应用的兼容性。最后,我们通过多种指标评估我们提出的模型,以评估其准确性、效率及在辅助和康复技术中部署的适用性。

B. 数据集

开源的 BASAN 数据集 [?] 包括来自 22 名年龄在 18 岁及以上的男性受试者下肢的四通道表面肌电图(sEMG)信号和一通道膝角度数据,这些数据是在三种下肢活动中记录的:在平地上行走(步态)、膝关节屈曲站立和膝关节伸展坐下。该数据集包括 11 名健康个体和 11 名膝盖异常者——其中包括 6 名前交叉韧带(ACL)损伤者、4 名半月板损伤者和 1 名坐骨神经损伤者。sEMG 信号使用 Datalog MWX8 记录,而膝关节角度数据使用 SG150B 测角仪收集。数据以 1000 Hz 采样,14 位分辨率。在采集后,sEMG 信号经过 20 Hz 到 460 Hz 的带通滤波。四个 sEMG 电极放置在膝关节屈伸涉及的关键肌肉上:股内侧肌(VM)、半腱肌(ST)、股二头肌(BF)和股直肌(RF)。在实验中,选择了健康个体的左腿以及膝盖异常者的患肢。需要强调的是,BASAN 数据集不包括过渡阶段数据,例如从坐到站立的动作。

在这项工作中,我们将数据预处理限制为在数据集中已经执行的步骤,如第 II-B 节所述。唯一的额外处理是将数据适当地构建为神经网络的输入。为了确保适当的泛化能力,我们采用留一验证法进行数据集划分。从 22 名参与者中,我们分配 9 名健康和 9 名异常个体用于训练,1 名健康和 1 名异常个体用于验证,1 名健康和 1 名异常个体用于测试。接下来,为了将数据格式化为神经网络输入,我们使用一个 256 毫秒的滑动窗口(对应于以 1000 Hz 采样率采样的 256 个数据样本),以及一个 64 毫秒的重叠,如图 ?? 所示。最后,训练数据被打乱以防止对特定个体或群体产生偏见。

所提出的神经网络以一个输入层开始,该层接收四个输入,分别对应四个表面肌电图(sEMG)通道,每个通道的序列长度为 256。随后,这些输入经过三个连续的卷积层,每个卷积层之后都有一个最大池化层。第一个卷积层使用

5×5 的卷积核和 16 个通道,第二个使用 3×3 的卷积核和 8 个通道,第三个使用 3×3 的卷积核和 4 个通道。在这三个阶段中,最大池化操作使用窗口大小为 2。特征提取后,每个通道的输出被展平成一个一维信号。这四个一维信号然后经过两个层,转换为两个独立的一维信号。这两个信号被连接形成一个单独的一维表示。所得信号然后通过一个包含 100 个神经元的全连接层处理,最后是一个包含三个神经元的全连接层负责分类。

本研究中采用的注意力机制旨在增强从每个输入窗口内各个单独的 sEMG 通道中提取特征。通过专注于信号中最具信息量的方面,注意力机制使网络能够强调对准确分类有贡献的相关模式。该机制与一个连接层结合使用,连接层充当一个直接的特征通路,使网络能够有效地保留和结合低级和高级特征。这些组件的协同作用改善了网络向后续层提供丰富和具有辨识特征的能力,从而提高分类性能。

具体来说,实现了一个 Bahdanau 注意机制,其中注意权重是通过一个专门的前馈神经网络来学习的。这使得模型能够动态调整其对输入特征的关注,学习为每个分类任务优先考虑最相关的时间模式。

为了防止过拟合,在第一个全连接层之前和最后一个全连接层之后应用了丢弃率为 0.5 的 dropout 层。网络中使用了 ReLU 激活函数,但 AM 层采用了 tanh 激活函数,而在最后一个全连接层中,使用了 softmax 激活函数进行分类。所提议网络的整体结构如图 ?? 所示。这种层的组合使得网络只有 62,876 个参数。

III. 结果与讨论

网络训练了 50 个周期,训练和验证损失结果如图 ?? (a) 所示。在训练和验证曲线中观察到的趋势表明,所提出的深度神经网络(DNN)在训练过程中成功地学习到了适当的映射,没有明显的欠拟合或过拟合迹象。在整个学习过程中,训练和验证准确率也观察到了类似的趋势,如图 ?? (b) 所示。在 50 个周期结束时,训练和验证准确率达到约 85%,进一步确认了模型良好的泛化能力。

为了评估模型在训练集之外的表现,该网络在测试对象中进行了验证。每个运动类别的接收者操作特征(ROC)曲线如图 ?? (c) 所示。曲线下面积(AUC)值分别为 0.96、0.95 和 0.95,显示出强大的分类性能。较高的 AUC 值接近 1.0,表示更好的分类置信度,这表明所提出的 DNN 能够有效区分不同的运动模式,同时将误报率降至最低。

为了更好地理解错误分类,我们还分析了混淆矩阵,如图 ?? (d) 所示。在这个矩阵中,对角元素表示正确分类的实例,而非对角元素表示错误分类的案例。经过分析结果,可以发现运动类别 1 在大多数情况下都被正确分类;然而,在 1288 个案例中,有 48 个被错误分类为运动类别 2,146 个被错误分类为运动类别 3。同样地,对于运动类别 2,在 910 个案例中,有 54 个被错误分类为运动类别 1,94 个被错误分类为运动类别 3。最后,对于运动类别 3,在 1016 个案例中,有 60 个被错误分类为运动类别 1,68 个被错误分类为运动类别 2。这些错误分类表明运动模式在 sEMG 信号空间中的特征存在重叠,导致分类时偶尔出现模糊性。这种重叠突显出仅仅凭借原始 sEMG 信号来区分某些运动的内在复杂性。

在测试对象上,所提出的模型实现了 85.38% 的总体准确率,表现出在区分三种运动类别时的强大性能。每种运动类别的分类结果总结在表格 ?? 中。基于 ?? 的类别评估显示,类别 1 具有最高的精确性(90.56%),表明对这种运动的预测非常可靠,误报极少。然而,其召回率(84.94%)

表明该类别的某些实例被误分类为其他运动。类别 2 的表现较为平衡,精确性为 86.79%,召回率为 83.74%,突显了模型在该类别中的良好泛化能力。相比之下,类别 3 具有最低的精确性 (78.72%),但具有最高的召回率 (87.40%),这意味着虽然模型能有效识别大多数类别 3 的运动,但更倾向于将其他运动误分类为类别 3。

从 F1-分数中可以得出类似的结果,该分数提供了对精确率和召回率的平衡测量。F1-分数针对每个类别进行计算,以评估整体分类性能。如表 ?? 所示,类别 1 获得了最高的 F1-分数 (87.66 %),这表明模型能够有效地检测和分类该运动,误报和漏报都很少。类别 2 的 F1-分数略低 (85.23 %),表明虽然该模型表现良好,但在该类别中精确率和召回率之间存在轻微的权衡。类别 3 的 F1-分数最低 (82.84 %),主要由于其精确率 (78.72 %) 低于召回率 (87.40 %)。总体而言,这些结果突出了在所有运动类别中的良好分类性能。

这种分类挑战可能源于类别 1 和 2 与类别 3 之间的根本差异,因为类别 3 对应于行走,它涉及更多的肌肉激活,从而导致更大的 sEMG 信号变异性。这些发现表明,尽管模型整体表现良好,但未来的改进可以关注于减少类别 3 中的误报,可能通过额外的特征提取技术或类别特定的训练策略来增强模型的区分能力。

为了进一步评估分类性能,我们计算了未见测试主体的平衡精度,其值为 85.36 %。平衡精度通过对所有类别的召回率进行平均来解决类别不平衡的问题,确保模型在不同运动类型上表现一致。该结果与总体精度 (85.38 %) 非常接近,表明模型并未偏向任何特定类别。

此外,为了支持我们设计轻量级网络的假设,我们通过测量推理时间来评估计算复杂性,即模型进行单次预测所需的时间。所提出的模型实现了每个样本平均推理时间小于 1 毫秒,展示了其适用于实时应用。这种效率使得该模型非常适合部署在假肢控制或康复系统中,在这些系统中,快速和可靠的运动分类是至关重要的。

与 [?] 中提出的模型相比,我们的网络表现出较低的准确性;然而,这一差异可以归因于几个因素。其中一个主要区别是我们研究中采用的留一法验证策略,这增强了模型的泛化能力。相比之下,[?] 将数据集整体划分为训练集、验证集和测试集,这意味着网络在训练期间可能部分见过验证和测试对象及其相关数据。这可能导致他们结果中更高的分类准确性。虽然将数据集整体处理是一个常见做法,但留一法确保了更稳健的评估,通过防止数据泄漏来获得更可泛化的结果。

此外,我们的方法不包含任何预处理技术,而 [?] 则利用了小波去噪。虽然小波去噪可以提高信号质量,但它也引入了高计算成本,使得网络在实时应用中不太可行。去噪所需的额外处理会导致延迟,这是实时假肢控制和康复系统中的一个关键限制。相比之下,我们的轻量级模型旨在优先考虑效率,以确保最小的计算开销和更快的响应时间,适合实际应用。

[?] 中提出的分类方法与我们的方法有若干相似之处,使其成为一个相关的基准用于比较。尽管他们的方法中也包含了小波去噪,但其数据分区策略与我们相似,允许通过对我们提出的方法进行一些小修改直接进行性能比较。一个关键的区别是 [?] 使用了一种基于多数投票的分类方法。在这种方法中,四个一维 CNN 网络并行运行,每个网络产生一个独立的分类结果,并通过投票机制确定最终结果。虽然这种技术提高了可靠性,但我们选择不在我们提出的 DNN 中使用它,因为其计算成本较高,特别是在

实时应用中,同时运行四个 DNN 网络会显著增加处理时间和资源消耗。

表 ?? 展示了在 [?] 中所提议网络的分类准确度,将多数投票法的准确度与我们不进行去噪的留一法进行比较。结果表明,我们提出的网络在保持较少参数数量的同时,性能优于基于多数投票的网络,突出了其计算效率。

为了进一步分析小波去噪的影响,我们对网络应用了 [?] 中描述的相同去噪技术。分类结果,包括那些使用去噪进行留一法验证的结果,汇总在表 ?? 中。尽管网络参数数量保持不变,但由于噪声降低过程,所有模型的准确性都得到了提高。基于多数投票的方法实现了 84.25 % 的验证准确率和 81.89 % 的测试准确率,而我们提出的网络结合小波去噪和留一法验证,实现了 87.67 % 的验证准确率和 86.39 % 的测试准确率。这些结果表明,我们的模型不仅达到了更高的准确率,还保持了较低的计算成本,使其更适合实时应用。

Tokas 等人最近提出了一种混合深度集成学习模型,该模型结合了卷积神经网络 (CNN) 和长短期记忆网络 (LSTM) 架构,以识别来自多通道表面肌电图 (sEMG) 信号的下肢活动 (坐着、站着、走路)。利用与我们的工作相同的 UCI sEMG 数据集,他们的方法在健康受试者中实现了令人印象深刻的分类准确率——分别为 99.3 %、98.3 % 和 98.8 %,用于行走、站立和坐着活动;在有病理的受试者中,分别为 99.0 %、98.1 % 和 98.2 %。然而,这些结果来源于 3 折交叉验证,而非留一受试者验证策略,后者更适合用于评估受试者无关情境下的泛化能力。此外,虽然作者没有报告推理时间,但其模型包括超过 141,000 个参数,大小是我们模型 (62,876 个参数) 的两倍多,暗示着一个更密集的架构,可能需要更高的计算量,这可能会阻碍其在边缘设备上实时部署。此外,他们的方法依赖于通过自适应合成采样 (ADASYN) 外推信号来平衡类别分布。尽管在离线环境中效果良好,此类合成过采样技术通常不适用于实时系统,实时系统中的数据不能以人工方式增强。因此,尽管离线表现强劲,该模型在实时和嵌入式环境中的适用性仍未得到验证。

IV. 结论

本研究表明,深度神经网络 (DNNs) 可以作为机器人系统中的上层控制器来实施。然而,要完全实现这一目标仍存在几个挑战。正如我们的结果所观察到的,类别检测中的不确定性和误分类仍然存在,这在现实场景中可能导致机器人执行不期望的动作。为降低这些风险,开发能够有效检测和解决误分类问题的鲁棒算法是至关重要的,以确保机器人决策的更高可靠性。

除了分类准确性之外,一个重要因素在之前研究中很大程度上被忽视了,包括我们的研究:人类感知和与这些高级控制器的交互作用。需要进一步调查人类用户如何感知和响应这些控制器,即使它们在当前状态中仍存在不完美之处。同时,从用户的角度评估这种智能控制机制是否提升了机器人系统的整体性能和可用性也至关重要。

该研究的一个可能起点可以是以被动模式测试上层控制器,允许机器人系统在无需直接影响控制决策的情况下运行。这种方法有助于探索人类与机器人之间的互动动力学,深入了解用户在实际场景中如何适应和从此类人工智能驱动的辅助中受益。

综上所述,本文介绍了一种基于注意力的深度神经网络,仅包含 62,876 个参数,旨在对下肢 sEMG 信号进行分类。所提出的模型在验证集 (新受试者) 上实现了 86.74 % 的

准确率，在测试集的留一检验条件下实现了 85.38 % 的准确率，仅需 50 个训练周期且不需要预处理。这些结果是在一个真实场景中获得的，验证集和测试集均由之前未见过的受试者构成，确保模型在真正新的数据上进行评估。

此外，即使在数据预处理中加入去噪滤波器，我们的模型在文献中仍优于以往的方法。这些发现突出了我们方法的效率和稳健性，使其成为机器人控制、康复和假肢系统中实时应用的有希望的候选方案。

V. 未来工作

虽然这项研究展示了对 sEMG 信号分类的有希望结果，但未来研究中仍有几个关键方面需要探索。其中一个重要研究领域是过滤对分类准确性的影响。由于过滤已经被证明能够提高准确性，因此需要进行更深入的分析以理解其对数据的影响，并开发实时可实施的过滤技术，以改善信号质量而不引入延迟。

另一个改进的途径在于通过超参数调整方法优化网络架构，这可以进一步提升模型的性能和计算效率。此外，在此类高级控制器的实际应用中，一个关键挑战是管理运动类别之间的转变，这一方面由于缺乏转变阶段的数据而未在本研究中解决。未来的研究应重点获取和分析转变阶段的数据，以确保在实际应用中的控制更加流畅和自然。