

基于图持续学习的时序数据分析^①

董次浩, 陈雷鸣, 黄子凌, 朱宜昌, 仇家康, 刘尚儒

(中国石油大学(华东), 青岛 266580)

通信作者: 董次浩, E-mail: dch142857@163.com



摘要: 随着可穿戴设备大规模进入生活, 基于动作传感器产生的时序数据来人体行为识别已成为该领域的研究热点. 然而目前的方法无法发现多个传感器数据在时空中相互作用的关系. 此外, 传统神经网络在学习新任务时, 由于学习的新任务参数会覆盖掉旧任务参数, 这会引起“灾难性遗忘”问题. 为解决这两个问题, 本文提出了一种基于图注意力网络与生成式回放持续学习机制融合方法的人体行为识别算法. 该算法通过卷积神经网络与图注意力网络提取时序特征, 使得模型能够同时关注时间与空间特征, 同时, 采用了基于生成式数据重放策略的情景记忆持续学习方法, 通过条件变分自编码器记忆历史数据分布来解决灾难性遗忘问题. 最后, 通过在多个公开数据集上与不同的基线算法对比, 实验结果表明本文所提算法可以在取得较高的准确率的同时, 缓解灾难性遗忘问题.

关键词: 图注意力网络; 可穿戴设备; 运动检测; 持续学习; 条件变分自编码器

引用格式: 董次浩, 陈雷鸣, 黄子凌, 朱宜昌, 仇家康, 刘尚儒. 基于图持续学习的时序数据分析. 计算机系统应用, 2024, 33(2): 188-197. <http://www.c-s-a.org.cn/1003-3254/9408.html>

Time Series Data Analysis Based on Graph Continuous Learning

DONG Ci-Hao, CHEN Lei-Ming, HUANG Zi-Ling, ZHU Yi-Chang, QIU Jia-Kang, LIU Shang-Ru

(China University of Petroleum, Qingdao 266580, China)

Abstract: With wearable devices entering life on a large scale, human behavior recognition based on temporal data generated by motion sensors has become a research hotspot in this field. However, the current methods cannot find the relationship between multiple sensor data in time and space. In addition, when the traditional neural network learns a new task, the new task parameters will overwrite the old task parameters, causing catastrophic forgetting problems. To this end, this study proposes a human behavior recognition algorithm based on the fusion method of graph attention network and generative playback continuous learning mechanism. The algorithm extracts temporal features through convolutional neural network and graph attention network, enabling the model to focus on temporal and spatial features at the same time. In addition, the algorithm adopts an episodic memory continuous learning method based on a generative data replay strategy, which remembers historical data distributions by conditional variational autoencoders, to address the catastrophic forgetting problem. Finally, compared with different baseline algorithms on multiple public datasets, the experimental results show that the proposed algorithm can achieve higher accuracy while mitigating the catastrophic forgetting problem more effectively.

Key words: graph attention network; wearable device; motion detection; continual learning; conditional variational autoencoder (CVAE)

① 收稿时间: 2023-08-24; 修改时间: 2023-09-26; 采用时间: 2023-10-07; csa 在线出版时间: 2023-12-25
CNKI 网络首发时间: 2023-12-27

作为当前的热点研究方向,基于传感器数据的人体动作分析技术广泛应用于医疗监测,体育赛事分析,自动驾驶和影视特效等领域.传统的传感器设备往往只能对当前记录的数据进行分析判别,而对历史数据的利用则较为粗糙,且流行的神经网络模型往往做不到同时利用时间和空间结构信息对数据进行处理.在时序数据的分类处理研究方面,目前主流的研究方法包括循环神经网络(recurrent neural network, RNN)^[1],及其为基础扩展与演化的模型,如长短时记忆模型(long short-term memory, LSTM)^[2],门控单元模型(gate recurrent unit, GRU)^[3]等.这些方法的共同优势在于能够充分利用时间序列的特征建模,但却忽略了重要的空间结构信息.此外,在真实环境下数据类型可能会发生动态变化,如生产环境中的新故障,医疗场景中的疾病的新分类等,而神经网络基于梯度对模型更新时,则会因为新知识对旧知识的干扰甚至覆盖而出现灾难性的遗忘问题,导致模型性能骤降,如图1所示.而将新旧任务数据合并,利用全部数据对重新训练模型,则会造成时间与计算资源的浪费.

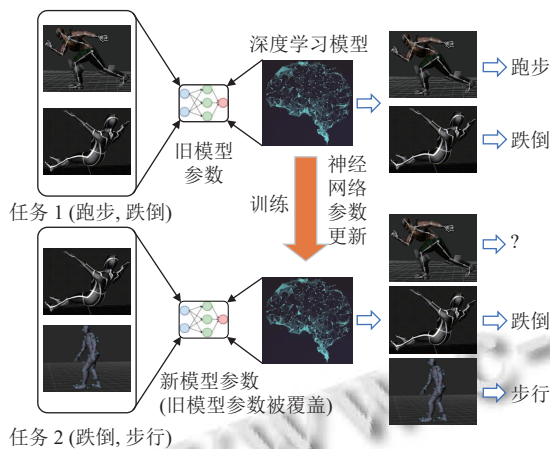


图1 深度学习中的灾难性的遗忘问题

针对以上两个问题,本文提出了图注意力网络与生成式回放机制融合方法(generated memory replay graph attention network, GMR-GAT).首先,利用卷积神经网络(convolutional neural networks, CNN)^[4],对传感器内局部时空信息进行特征提取,之后利用图注意力网络(graph attention networks, GAT)^[5]对传感器间的空间及时间结构信息进行处理,其中注意力机制可以使模型能够关注更重要的时间或空间节点,进而提高模型的性能^[6].此外,引入了基于条件变分自编码器(conditional variational autoencoders, CVAE)^[7]作为深度生成模型,以在不存储原始数据的条件下,针对样例数据分布进行记忆,通过情景记忆回放辅助训练的方式缓解灾难性遗忘问题.

目前,图神经网络已广泛用于解决图结构化数据的问题^[8].Gori等^[9]使用了RNN以提取节点信息并学习图节点的标签,图神经网络这一概念自此诞生.之后文献^[10]提出图卷积网络(graph convolutional network, GCN),提出利用CNN对图结构的数据进行建模,使得模型能够提取中心节点和邻居节点的特征和标签信息,并将其输入到CNN中,这令GCN具备了提取多尺度信息的能力,从而实现对图结构信息和属性信息的有效利用^[11,12].注意力机制通过选择特定的输入,使神经网络更关注那些与学习任务关联度更高的信息^[13].文献^[5]首次提出将注意力机制应用到GNN,二者结合后形成了图注意力网络(graph attention network, GAT).但只利用常规图神经网络无法克服深度学习出现的灾难性遗忘问题.

1 相关工作

持续学习指的是模型在保留先前所学知识的同时,不断学习新知识,并避免对旧任务的遗忘与性能骤降.使人工智能系统具备持续学习能力是当今智能学习领域面临的重要挑战^[14,15].目前主流的持续学习方法包括以可塑权重巩固(elastic weight consolidation, EWC)^[16],突触智能(synaptic intelligence, SI)^[17],无遗忘学习(learning without forgetting, LwF)^[18],记忆感知突触(memory aware synapses, MAS)^[19]为代表的通过对模型实施正则化约束以避免遗忘的正则化方法和深度生成重放(deep generative replay, DGR)^[20],经验回放(experience replay, ER)^[21],梯度情景记忆(gradient episodic memory, GEM)^[22]及平均梯度情景记忆(averaged gradient episodic memory, A-GEM)^[23]为代表的重复学习以避免遗忘的情景记忆回放方法等.然而,正则化方法下,模型的大小会随任务数量的增加而迅速增加,正则化效果则会逐渐下降.情景记忆回放方法在减轻灾难性遗忘方面具有巨大优势,然而,该方法需要长期保存真实样例,不利于数据安全保护.

自编码器(autoencoder, AE)是由Rumelhart等^[24]提出的一种无监督学习算法,该算法的目的之一是对输入其中的数据进行压缩,并且能够复现输出. Kingma

自编码器(autoencoder, AE)是由Rumelhart等^[24]提出的一种无监督学习算法,该算法的目的之一是对输入其中的数据进行压缩,并且能够复现输出. Kingma

等^[25]提出的变分自编码器 (variational autoencoder, VAE) 是一类应用较广的自编码器, 也是一类重要的生成模型. 变分自编码器能够以特定分布的随机样本作为输入, 生成相应的仿真样例. 在此基础上, Sohn 等^[7]改进并提出了条件变分自编码器, 使模型能够根据特定条件, 如分类标签等, 合成对应的数据^[26].

2 算法设计

为实现传感器时空结构数据的有效利用, 同时实现不存储原始数据条件下的持续学习功能, 本文提出

了一种图注意力网络与生成式回放机制融合方法, 首次采用了条件变分自编码器作为深度生成重放中的深度生成模型, 结合了深度生成重放与经验回放两种记忆持续学习方法的优点, 避免了经验回放方法中, 训练方需要长期保留数据提供方提供的真实数据而可能造成的安全与隐私问题, 同时, 利用经验回放方法中的等概率记忆采样技巧, 稳定了生成数据类别的分布, 使最终用于回放的生成数据更贴近原始分布. 算法的输入为不同时段的任务, 不同任务包含不同类别的多维时序数据. 简化的算法流程如图2所示.

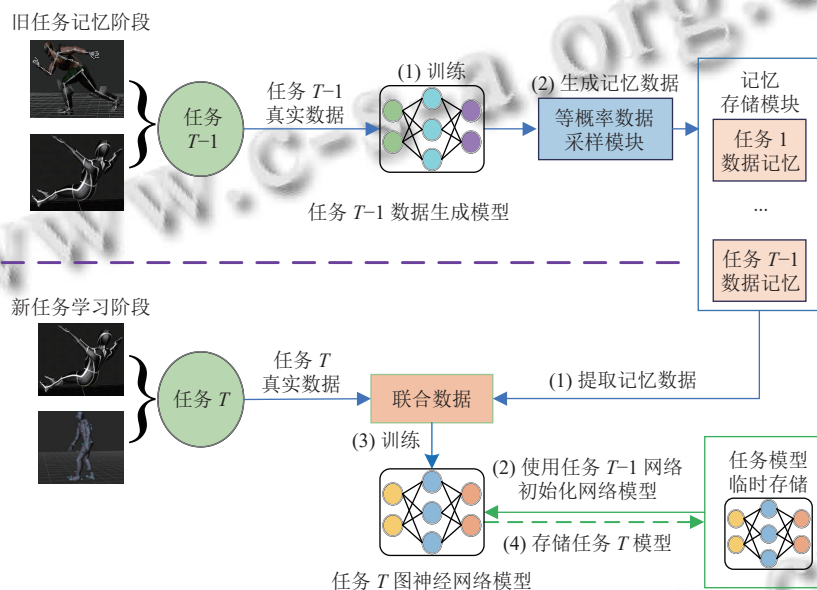


图2 GMR-GAT 算法流程

旧任务记忆阶段: 该阶段负责完成生成模型的训练和旧任务的记忆, 为新任务的学习做准备. 主要分为2个流程, 详细的流程描述如下.

(1) 训练网络模型: 使用 $T-1$ 时刻任务的数据分别训练 $T-1$ 时刻的网络模型和 $T-1$ 时刻的数据生成模型.

(2) 生成记忆数据: 使用 $T-1$ 时刻任务训练的数据生成网络进行仿真数据输出, 同时将生成数据通过等概率数据采样存入记忆存储模块.

新任务学习阶段: 该阶段主要完成任务 T 的图神经网络模型训练, 同时缓解对任务 $T-1$ 的遗忘. 主要分为4个流程, 详细的流程描述如下.

(1) 提取记忆数据: 提取记忆存储模块中记录的记忆数据, 将记忆数据与新任务数据合并为联合数据.

(2) 初始化网络模型: 从任务模型存储模块中拉取旧任务的模型参数完成本次任务的网络初始化.

(3) 训练网络模型: 使用上一步初始化的网络和联合数据集进行本轮任务的训练.

(4) 存储模型: 将针对任务 T 训练后的模型入模型存储模块.

假设整个数据集被划分为 T 个任务, 重复执行以上两个阶段直到整个训练结束, 最后获得 T 任务结束后的网络模型, 使用该模型在测试集上进行测试.

本文所提出的改进后的图注意力网络, 在获取训练数据后, 首先按照传感器的时间与空间结构进行分割处理, 分割后的每一部分包含单个传感器内部一段时间的时序数据, 利用卷积神经网络对每个传感器的局部数据进行特征提取后, 形成多条传感器特征数据, 其数量与传感器数量, 分割时段数量有关. 将每条数据视为一个节点, 构建为全连接结构的图, 节点间的边代表传感器间的时空联系. 随后利用图注意力网络对形

成的图结构数据进行处理,并输出类别信息.

2.1 深度生成模型结构设计

由于将真实数据长期存放于训练方会存在安全隐私风险,文献[20]提出可以利用深度模型记忆数据分布,再利用生成的数据进行回放的思路.受该理论的启发,本文构造了数据生成回放模块,提出了一种利用CVAE生成数据进行回放式记忆的处理框架.

变分自编码器由两部分组成:将输入数据压缩为隐空间(latent space)中编码数据的编码器(encoder)和用于解压缩的解码器(decoder).解码器应根据隐空间采样隐变量数据 z ,输出准确表示原始数据分布的样本,因此需要对未知概率分布函数进行建模.从隐空间变量分布构建原始数据分布 $p(X)$ 的过程为:

$$p(X) = \sum_Z p(X|Z)p(Z) \quad (1)$$

其中, $p(Z)$ 对应为事先指定的隐变量分布,通常为标准正态分布. $p(X|Z)$ 对应于解码器所对应的由编码变量到解码变量的分布,假设服从正态分布.同样, $p(Z|X)$ 为编码器所对应的由原始变量到编码变量的分布,且不难证明, $p(Z|X)$ 也满足正态分布.

根据贝叶斯公式,可知:

$$p(Z|X) = \frac{p(X|Z)p(Z)}{p(X)} = \frac{p(Z)p(X|Z)}{\int p(X|z)p(z)dz} \quad (2)$$

然而,在信息受限的情况下,计算该公式的代价极大.故采用变分贝叶斯方法,借助另一个分布 $q(Z|X)$ 来尽可能逼近 $p(Z|X)$.

而CVAE的输入除样本数据 x 外,还包括标签数据 y .根据上述理论,构建CVAE模型如图3所示,其运行流程如下.

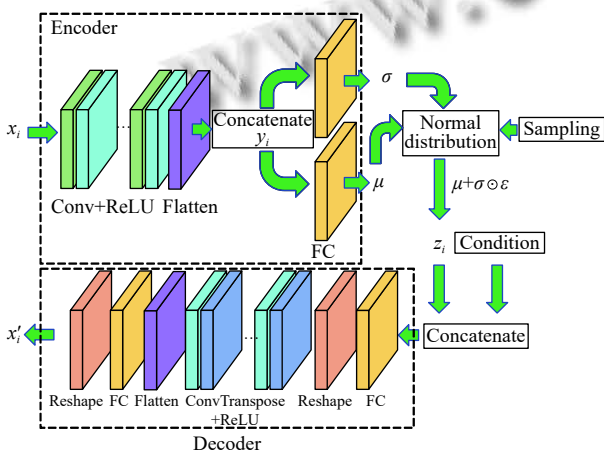


图3 CVAE结构示意图

(1) 将数据 x_i, y_i 处理并拼接后输入编码器,通过神经网络得到隐变量 z 服从的近似后验分布 $q_\phi(z|x, y)$ 的参数,对正态分布的参数 σ^2 和 μ 进行拟合.

(2) 根据编码器得到的参数 σ^2 和 μ ,从标准正态分布中采样得到 ϵ ,令 $z_i = \mu + \sigma \odot \epsilon$,其中 \odot 代表逐元素相乘, z_i 服从 σ^2 和 μ 对应的正态分布.

(3) 将 z_i 与指定的生成类别 c (训练中为 y_i)拼接后输入解码器,同样通过神经网络拟合 $p(x|z)$,并根据该分布生成均值数据作为生成结果 x'_i 并输出.

针对编码器的拟合过程,利用KL散度损失函数,衡量编码器拟合的分布与标准正态分布之间的差异.针对解码器的拟合过程,利用重构损失衡量输入样本与合成数据之间的差异,本文中使用了MSE损失函数.如式(3)所示:

$$loss = MSE(x_i, x'_i) + KL(N(\mu, \sigma^2), N(0, 1)) \quad (3)$$

其中,第2项的计算方式为:

$$KL(N(\mu, \sigma^2), N(0, 1)) = \frac{\mu^2 + \sigma^2 - 2\log(\sigma) - 1}{2} \quad (4)$$

2.2 等概率记忆采样

为保证数据集中的每个样本被记忆的概率相同,本文采用了一种等概率样本选择方法.

设每个样本数据被保存在记忆中的概率为 P ,记忆存储大小为 M ,已处理的数据量为 n ,则对不同条件下的样本有:

$$P = \begin{cases} 1, & n \leq M \\ M/n, & n > M \text{ 且未被处理} \\ (n-1)/n, & n > M \text{ 且已存储} \end{cases} \quad (5)$$

由上述公式,可进行如下验证,设 $P_{t,t'}$ 为任意 t 时刻未被处理的数据在其后的 t' 时刻中存在于记忆中的概率,可得:

$$P_{t,t'} = \frac{M}{n_t} \times \frac{n_t}{n_{t+1}} \times \dots \times \frac{n_{t'-2}}{n_{t'-1}} \times \frac{n_{t'-1}}{n_{t'}} = \frac{M}{n_{t'}} \quad (6)$$

其中, n_t 代表 t 时刻处理的数据量.式(6)可证明 $P_{t,t'}$ 只与存储大小以及当前时刻的数据量有关,可以实现样本数据的等概率记忆采样.

2.3 图神经网络结构设计

可穿戴式运动传感器,能够采集并记录所佩戴部位运动时所产生的多维时序数据.对于收集到的多维时序数据,其特征维度以 F 表示,传感器数量记为 S .

并将来自单个传感器 t 时段内的 N 个连续数据点视为一个数据块进行, 在 t 时段内, 数据形式为 $x_t^s \in \mathbb{R}^{N \times F}$, 其中 $s = 1, 2, \dots, S$. 如图 4 所示. 而由于人体运动与时空结构密切相关, 可以假设 $x_{t-1}^1, \dots, x_{t-1}^s, x_t^1, \dots, x_t^s$ 之间存在潜在的关系, 并将其视为具有全连接关系的不同节点, 组成图结构数据. 此外, 从空间和时间的角度来看, 传感器之间联系的重要性也存在一定差异, 例如, 在跑步动作中, 当前时刻的左臂动作与上一时刻的右臂动作间的关联性可能与当前时刻的左臂动作, 上一时刻背部动作之间的关联性有所区别.

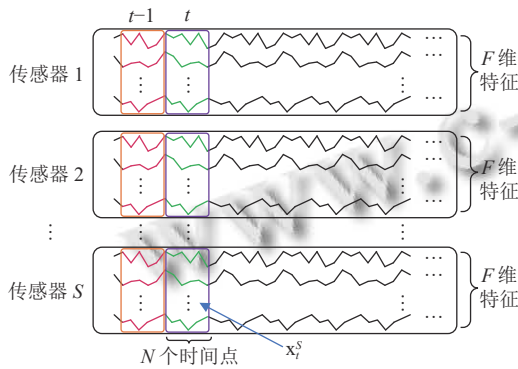


图 4 传感器时空结构

本文首先利用卷积神经网络对节点内局部数据进行特征提取, 将其转化为 F 维特征后, 将所有节点的特征组成全连接图, 作为图注意力网络的输入: $h = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_{P \times S}\}$, $\vec{h}_i \in \mathbb{R}^F$, 其中 P 为参与计算的时段数量, 图注意力网络的输出为新的节点特征集合 $h' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_{P \times S}\}$, $\vec{h}'_i \in \mathbb{R}^F$. 计算方法如下式所示:

$$\vec{h}'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} W \vec{h}_j \right) \quad (7)$$

其中, σ 为 ReLU 激活函数, N_i 为与节点 \vec{h}_i 在时间或空间上相邻的节点, W 为可学习参数, α_{ij} 为注意力系数, 计算方法如式 (8) 所示:

$$\alpha_{ij} = \frac{\exp \left(\text{LeakyReLU} \left(\vec{a}^T \left[W \vec{h}_i \| W \vec{h}_j \right] \right) \right)}{\sum_{k \in N_i} \exp \left(\text{LeakyReLU} \left(\vec{a}^T \left[W \vec{h}_i \| W \vec{h}_k \right] \right) \right)} \quad (8)$$

其中, \vec{a}^T 的作用是将高维特征映射为一个实数, 通过单层前馈神经网络实现.

最后通过全连接层, 采用 Softmax 对提取到的特征进行分类输出. 其整体结构其如图 5 所示.

3 实验分析

3.1 实验环境

本文中进行的实验均基于 Ubuntu 20.04 操作系统; 硬件方面, 所采用 GPU 为 Nvidia RTX 2080Ti, 显存大小为 11 GB, CPU 采用 Intel(R) Core i9-9900K @ 3.60 GHz $\times 8$, 运行内存大小为 32 GB; 软件方面, Python 版本为 3.8.10, PyTorch 为 1.12.0, CUDA 版本为 11.3.

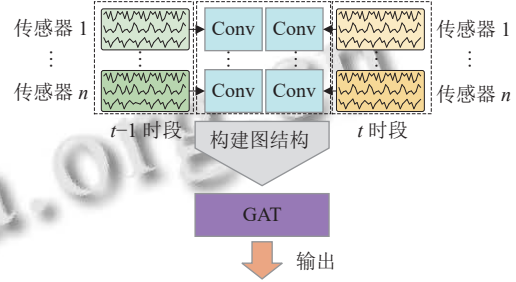


图 5 改进的图神经网络结构

3.1.1 数据集描述

如表 1 所示, 本文将所构建框架分别在 3 个不同的 UCI 公共数据集上进行训练和测试, 从各类别中采集 1000 条, 并按照 7:3 的比例划分测试集与训练集.

表 1 实验中使用数据集的统计数据

数据集	特征	类别	总量
DSADS	45	19	19000
MHEALTH	23	12	12000
REALDISP	117	33	33000

3 个数据集包括日常和活动数据集 (daily and sports activities data set, DSADS), 移动健康监测数据集 (mobile health, MHEALTH), 现实传感器位移数据集 (realistic sensor displacement, REALDISP).

DSADS 数据集由 45 维, 19 类不同活动的传感器时序数据组成. 其中传感器分布在身体的不同部位, 采集了包含各方向的加速度计, 陀螺仪, 磁强计等活动类型的数据, 活动类型包含上下楼梯, 跑步, 跳跃, 骑自行车等常见活动类型. MHEALTH 数据集中包括了 23 维, 12 类活动的时序数据. 23 维中, 有 21 维是分布在身体的不同部位的运动传感器数据, 2 维是心电信号数据, 活动类型包括弯腰, 屈膝, 躺下等. REALDISP 数据集包含了来自不同部位传感器的 117 维, 33 类活动的加速度计, 陀螺仪, 磁强计和四元数的时序数据, 活动类型包括跳绳, 前后跳跃, 旋转手臂等.

3.1.2 实验指标

为了更有效地评估持续学习的性能, 本文引入了

该领域流行的两种指标: 平均准确率^[22], 定义为 A_q , 以及平均遗忘率^[27], 定义为 F_q . 其中, 平均准确率评估了当前模型在所有任务上的表现, 数值越大则表现越好. 平均遗忘率衡量了学习新任务而导致对旧任务的遗忘程度, 其数值越小代表遗忘程度越低.

平均准确率的计算公式为:

$$A_q = \frac{1}{q} \sum_{p=1}^q a_{q,p} \quad (9)$$

其中, $a_{q,p}$ 为模型学习完第 q 个任务后, 在第 p 个任务的测试集的准确率.

定义模型在完成任务 q 的训练之后, 对于第 p 个任务的遗忘程度的量化形式为遗忘率 f_p^q , 代表了任务 p 在持续学习过程中所取得的最好表现与在学习完 q 个任务后, 模型对任务 p 的表现的差异程度, 其计算公式为:

$$f_p^q = \max_{o \in \{1, \dots, q-1\}} a_{o,p} - a_{q,p} \quad (10)$$

通过模型对所有旧任务的遗忘率进行平均计算, 得到平均遗忘率的计算公式为:

$$F_q = \frac{1}{q-1} \sum_{p=1}^{q-1} f_p^q \quad (11)$$

3.2 实验设计

3.2.1 特征提取实验

为了验证本文所提的特征提取方法的有效性, 首先将所提出的算法在总数据集上与多种基线算法进行了对比. 本实验方法采用的优化方式为随机梯度下降法 (SGD), 学习率设置为 0.01, 训练批次大小为 10, 训练迭代次数为 50. 基线算法包括图卷积网络 (GCN) 与

长短期记忆神经网络 (LSTM).

将上述建模方法与本文所提出的建模方法进行实验对比. 在全类别训练集的条件下, 采集不同时间长度作为历史信息条件下进行实验, 以测试集准确率作为评价指标, 具体结果如表 2 所示.

表 2 不同采样长度下特征提取性能对比

方法	数据集	10长度	20长度	30长度	40长度	平均值
GCN	DSADS	0.9026	0.9211	0.9177	0.8858	0.9068
	MHEALTH	0.9079	0.8854	0.9204	0.8892	0.9007
	REALDISP	0.9189	0.9001	0.9064	0.8381	0.8909
LSTM	DSADS	0.9159	0.9294	0.9355	0.9371	0.9295
	MHEALTH	0.9388	0.9518	0.9431	0.9529	0.9467
	REALDISP	0.9283	0.9314	0.9385	0.8692	0.9169
GAT	DSADS	0.9459	0.9552	0.9578	0.9492	0.9520
	MHEALTH	0.9502	0.9796	0.9821	0.9791	0.9728
	REALDISP	0.9622	0.9674	0.9547	0.9683	0.9632
GAT*	DSADS	—	—	—	—	0.8961
	MHEALTH	—	—	—	—	0.8575
	REALDISP	—	—	—	—	0.8767

表 2 中的 GAT* 代表不引入历史数据, 即只利用当前时刻的数据进行训练的表现. 通过表中的结果可以看出, 较不采用注意力机制的 GCN, 忽略数据空间结构的 LSTM 和忽略数据的时间属性的情况 GAT*, 采用本文所提的建模方法在不同数据集上均具有更好的表现, 验证了其有效性, 即能够更有效地处理包含空间结构的多维时序数据.

另外, 使用本文所提出的生成模块, 在不同时间采样长度下对所用数据集的部分生成结果进行了可视化, 如图 6-图 8 所示, 左侧为原始数据, 右侧为仿真数据. 纵坐标代表采样长度, 横坐标代表数据维度, 颜色代表数据大小. 可以看出, 采用条件变分自编码器可以有效地实现对真实数据的仿真合成.

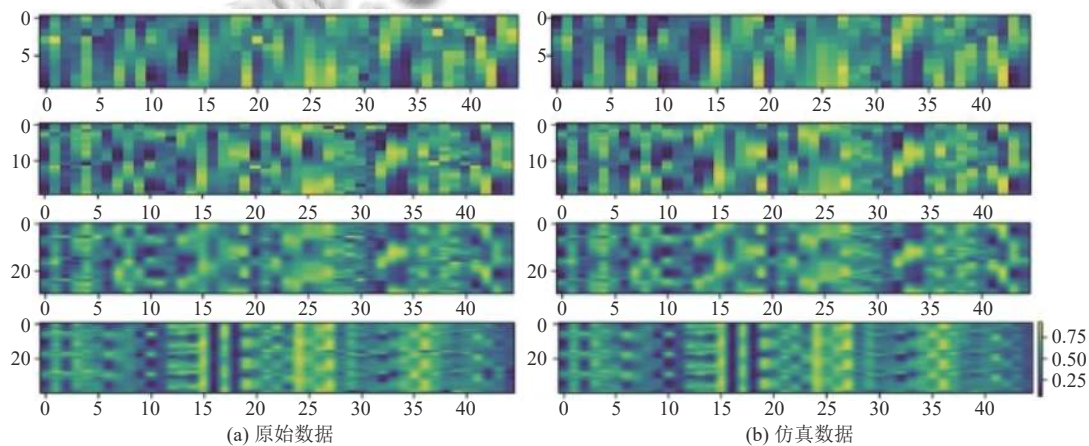


图 6 DSADS 数据集仿真效果 (纵坐标代表采样长度, 横坐标代表数据维度)

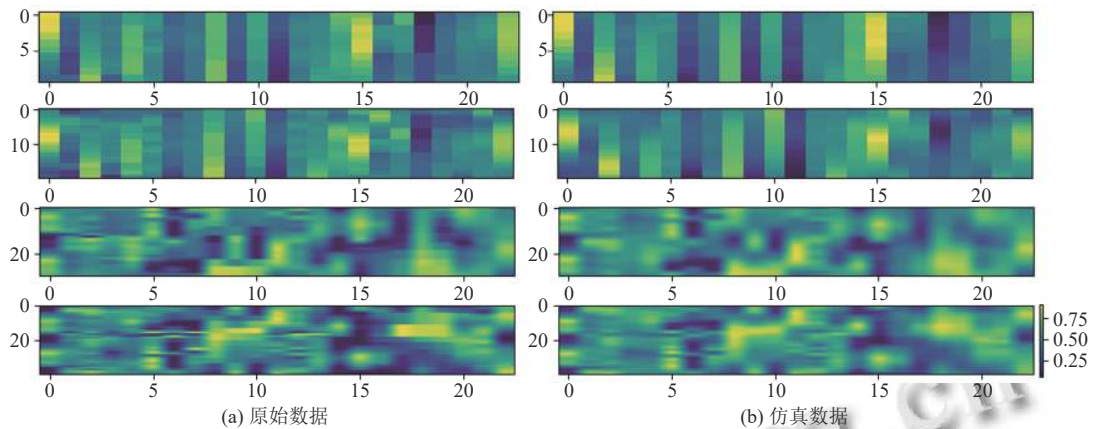


图7 MHEALTH数据集仿真效果(纵坐标代表采样长度,横坐标代表数据维度)

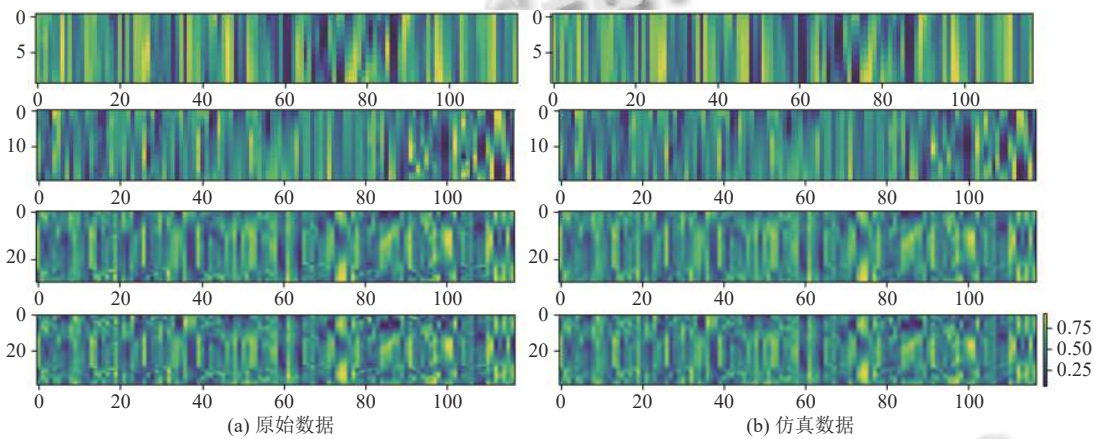


图8 REALDISP数据集仿真效果(纵坐标代表采样长度,横坐标代表数据维度)

3.2.2 持续学习实验

本文将所提算法与多种持续学习方法进行了比较: (1) LwF^[18], 利用知识蒸馏策略缓解对旧任务知识的遗忘. (2) EWC^[16], 通过限制新旧任务之间模型参数的变化, 实现对旧任务知识的记忆. (3) GEM^[22], 利用记忆情景模型限制参数更新方向, 避免针对旧任务的损失上升. (4) MAS^[19], 通过寻找对旧任务重要的参数, 并降低其变化幅度, 防止与旧任务相关的重要知识被覆盖.

在本文设置的持续学习环境中, 训练数据被划分为任务队列, 由包含不同数据的任务组成. 为模拟数据的持续到来, 早期的任务只包含较少类别的数据, 且一个任务内最多容纳3类数据. 设 c 代表数据类别, k 代表数据集的类别总数, 则有任务列表 T 的组成结构为 $\{[c_1], [c_1, c_2], [c_1, c_2, c_3], \dots, [c_{k-2}, c_{k-1}, c_k]\}$.

本文在 DSADS, MHEALTH 与 REALDISP 这3个数据集上进行了实验, 以平均准确率和平均遗忘率

作为评价指标. 实验结果的平均准确率如表3所示, 平均遗忘率如表4所示, 此外, 图9-图11展示了持续学习过程中各方法下平均准确率的变化过程, 其中 GMR-GAT 是本文所提出的方法.

表3 不同方法平均准确率比较

方法	DSADS	MHEALTH	REALDISP
LwF	0.6444	0.7552	0.5782
EWC	0.8514	0.9025	0.808
GEM	0.8297	0.8863	0.8816
MAS	0.7798	0.9251	0.8238
GMR-GAT	0.9341	0.9568	0.9222

表4 不同方法平均遗忘率比较

方法	DSADS	MHEALTH	REALDISP
LwF	0.3381	0.2504	0.3757
EWC	0.1236	0.093	0.135
GEM	0.112	0.0872	0.0831
MAS	0.1829	0.0691	0.1447
GMR-GAT	0.0351	0.0342	0.0441

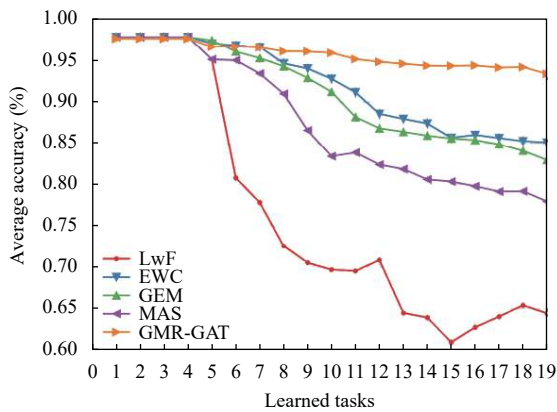


图9 DSADS数据集平均准确率变化过程

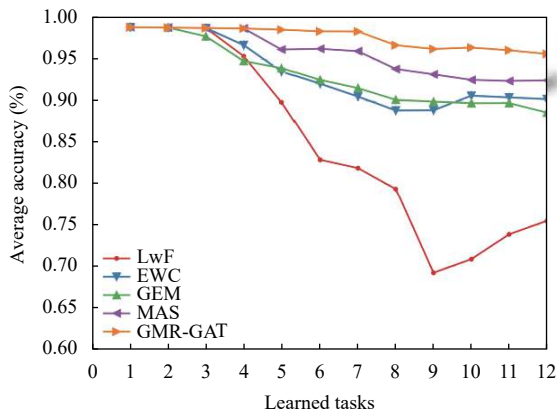


图10 MHEALTH数据集平均准确率变化过程

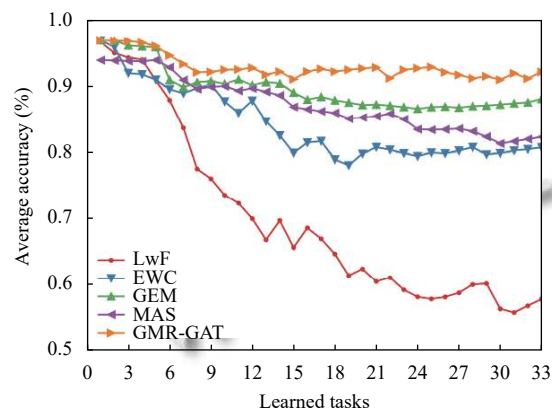


图11 REALDISP数据集平均准确率变化过程

可以看出, 本文提出的方法在3个数据集上具有较高的平均准确率和较低的平均遗忘率. 灾难性遗忘的根本原因是在学习新任务时, 神经网络的权重会被不断动态调整以优先适应当前任务的数据. 而本实验的结果证明, 通过回放情景记忆, 可以使模型在适应当前任务的同时, 维持一部分有利于旧任务的神经网络

参数, 通过这种方式所得到的模型能够在新任务与旧任务上均取得较好的表现, 实现了对灾难性遗忘的缓解. 而LwF在所有的数据集中的表现不如其他基线, 这表明LwF不能很好地适应前文所提的特征提取方法. 在训练新任务时限制重要模型参数变化的EWC和MAS方法, 在不同数据集上训练所得模型的表现较不稳定, 数据集的选取对测试结果存在较大影响. GEM在保持旧任务数据损失的同时, 限制模型参数的更新, 这使其在不同数据集上表现稳定, 且能取得较低的平均遗忘率, 但同时抑制了对新任务的学习, 导致其平均准确率的下降.

3.3 消融实验

为了验证情景记忆缓冲区大小对实验结果的影响, 本文使用了不同大小的记忆缓冲区的场景进行实验. 设缓冲区大小分别为平均每类20, 25, 30, 35, 40. 不同大小的记忆缓冲区下对平均准确率和平均遗忘率的影响分别如表5, 表6所示.

表5 忆缓冲区大小对平均准确率的影响

类平均空间	DSADS	MHEALTH	REALDISP
20	0.4838	0.3992	0.6615
25	0.9167	0.6333	0.8437
30	0.9341	0.855	0.9222
35	0.9408	0.9492	0.9387
40	0.9452	0.9568	0.9418

表6 缓冲区大小对平均遗忘率的影响

类平均空间	DSADS	MHEALTH	REALDISP
20	0.5435	0.655	0.3463
25	0.122	0.3995	0.152
30	0.0351	0.1577	0.0441
35	0.0295	0.1345	0.0429
40	0.0264	0.0342	0.0422

通过表中结果可以看出, 在记忆缓冲区较小时, 模型的持续学习效果较差, 持续学习效果随记忆缓冲区增大而变强, 但当缓冲区增大到一定程度时, 对持续学习效果的增强效果逐渐不明显.

为了验证等概率采样算法的有效性, 本文对采用不同方法构建缓冲区的情景记忆方式进行了实验测试, 其中, 当缓冲区满后, 随机采样策略1将缓冲区内的数据与当前数据合并后进行随机筛选丢弃, 最终缓冲区内中数据量满足限制要求, 随机采样策略2在数据到来前先对缓冲区内数据进行随机筛选丢弃, 数据到来后进入缓冲区, 也使缓冲区内中数据量满足限制要求.

不同策略下平均准确率表现如表7所示,平均遗忘率表现如表8所示。

表7 不同策略下的平均准确率

采样方法	DSADS	MHEALTH	REALDISP
等概率采样	0.9341	0.9568	0.9222
随机采样1	0.5342	0.4572	0.5542
随机采样2	0.3654	0.3597	0.4632

表8 不同策略下的平均遗忘率

采样方法	DSADS	MHEALTH	REALDISP
等概率采样	0.0351	0.0342	0.0441
随机采样1	0.4843	0.5868	0.7339
随机采样2	0.6641	0.6938	0.8464

实验结果表明,等概率记忆采样方法能够有效提高本算法的持续学习效果。

为了测试不同数据量环境下利用生成模块产生合成数据进行训练对模型效果的影响,分别设置了利用40%、70%和100%数据的训练环境,并对比了去除生成器模块,仅回放原始数据的策略。实验结果的平均准确率对比表现如表9所示,平均遗忘率对比表现如表10所示。

表9 不同数据量下的平均准确率对比

所用数据比例 (%)	数据集	无合成数据	合成数据
40	DSADS	0.6638	0.6818
	MHEALTH	0.6921	0.7192
	REALDISP	0.5793	0.6081
70	DSADS	0.8405	0.8418
	MHEALTH	0.8696	0.8817
	REALDISP	0.7959	0.7981
100	DSADS	0.9416	0.9341
	MHEALTH	0.9523	0.9568
	REALDISP	0.9254	0.9222

表10 不同数据量下的平均遗忘率对比

所用数据比例 (%)	数据集	无合成数据	合成数据
40	DSADS	0.1432	0.1349
	MHEALTH	0.2268	0.1932
	REALDISP	0.2142	0.1778
70	DSADS	0.0637	0.0549
	MHEALTH	0.0818	0.0691
	REALDISP	0.0991	0.091
100	DSADS	0.0378	0.0351
	MHEALTH	0.0394	0.0342
	REALDISP	0.0511	0.0441

根据实验结果可以看出,随着数据量的减少,使用合成数据的方式的平均准确率下降速度及平均遗忘率上升速度低于不使用的情况,总体上具有更好的效果,

能够在不影响模型效果的前提下,通过不断回放合成数据,缓解模型的灾难性遗忘问题,同时实现对原始数据的保护。

4 总结

针对当前主流多维时序数据分析算法难以有效利用可穿戴设备传感器数据的空间性与时间性结构特点的问题,及多维时序数据分析结束后,再对新类别数据进行训练出现的灾难性遗忘,主流算法需要重新对整体数据进行训练的问题,本文在图注意力网络的研究基础上,改进了采样与处理方式,提出了一种具有持续学习能力的算法框架。

首先利用卷积神经网络和对传感器数据中的局部时间和空间信息进行提取:之后利用图注意力网络对各传感器之间的时间和空间信息进行处理,通过注意力机制对不同的时空节点赋予不同权重以提升特征提取效果。经过以上两个步骤,能够同时提取到传感器的时间特征与空间特征,并通过与基线算法进行对比实验和消融实验,验证了本文所使用的特征提取模块的有效性。此外,设计了一种合成数据生成模块与对应的情景回放机制,通过利用真实数据,合成数据及情景记忆数据对模型进行训练,缓解了模型的灾难性遗忘问题。最后,将该模型在3种不同的可穿戴设备传感器数据集中进行验证,结果表明,与其他处理方式相比,本文提出的算法框架不仅具有更高的准确率,并且在持续学习方面的平均准确率与平均遗忘率这两种评价指标上均优于其他持续学习方法,在训练时间受限等特殊环境中具有一定的实用价值。由于目前采用的神经网络仍较为基础,未来考虑将CNN替换为其他更先进的神经网络模型。

参考文献

- Elman JL. Finding structure in time. *Cognitive Science*, 1990, 14(2): 179–211. [doi: 10.1207/s15516709cog1402_1]
- Graves A. Long short-term memory. *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Heidelberg: Springer, 2012. 37–45.
- Dey R, Salem FM. Gate-variants of gated recurrent unit (GRU) neural networks. *Proceedings of the 60th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. Boston: IEEE, 2017. 1597–1600.
- Gu JX, Wang ZH, Kuen J, *et al.* Recent advances in

- convolutional neural networks. *Pattern Recognition*, 2018, 77: 354–377. [doi: [10.1016/j.patcog.2017.10.013](https://doi.org/10.1016/j.patcog.2017.10.013)]
- 5 Velickovic P, Cucurull G, Casanova A, *et al.* Graph attention networks. *Proceedings of the 6th International Conference on Learning Representations*. Vancouver: OpenReview.net, 2018.
- 6 徐子路, 朱睿莎, 余敦辉, 等. 面向图注意力网络的突发热点事件联合抽取. *小型微型计算机系统*, 2023, 44(5): 902–909.
- 7 Sohn K, Yan XC, Lee H. Learning structured output representation using deep conditional generative models. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015. 3483–3491.
- 8 Zhou J, Cui GQ, Hu SD, *et al.* Graph neural networks: A review of methods and applications. *AI Open*, 2020, 1: 57–81. [doi: [10.1016/j.aiopen.2021.01.001](https://doi.org/10.1016/j.aiopen.2021.01.001)]
- 9 Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains. *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks*. Montreal: IEEE, 2005. 729–734.
- 10 Bruna J, Zaremba W, Szlam A, *et al.* Spectral networks and locally connected networks on graphs. *Proceedings of the 2nd International Conference on Learning Representations*. Banff: ICLR, 2014. 1–14.
- 11 位俊超, 陈春雨. 基于 SAT-GCN 的花样滑冰选手动作检测算法研究. *应用科技*, 2023, 50(1): 7–13. [doi: [10.11991/yykj.202204003](https://doi.org/10.11991/yykj.202204003)]
- 12 黄玉娇, 陈铭凯, 郑媛, 等. 基于弱化图卷积网络的文本分类. *计算机科学*, 2023, 50(6A): 220700039. [doi: [10.11896/jsjx.220700039](https://doi.org/10.11896/jsjx.220700039)]
- 13 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 14 Hassabis D, Kumaran D, Summerfield C, *et al.* Neuroscience-inspired artificial intelligence. *Neuron*, 2017, 95(2): 245–258. [doi: [10.1016/j.neuron.2017.06.011](https://doi.org/10.1016/j.neuron.2017.06.011)]
- 15 French RM. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 1999, 3(4): 128–135. [doi: [10.1016/S1364-6613\(99\)01294-2](https://doi.org/10.1016/S1364-6613(99)01294-2)]
- 16 Kirkpatrick J, Pascanu R, Rabinowitz N, *et al.* Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(13): 3521–3526.
- 17 Zenke F, Poole B, Ganguli S. Continual learning through synaptic intelligence. *Proceedings of the 34th International Conference on Machine Learning*. Sydney: PMLR, 2017. 3987–3995.
- 18 Li ZZ, Hoiem D. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(12): 2935–2947. [doi: [10.1109/TPAMI.2017.2773081](https://doi.org/10.1109/TPAMI.2017.2773081)]
- 19 Aljundi R, Babiloni F, Elhoseiny M, *et al.* Memory aware synapses: Learning what (not) to forget. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 144–161.
- 20 Shin H, Lee JK, Kim J, *et al.* Continual learning with deep generative replay. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 2994–3003.
- 21 Rolnick D, Ahuja A, Schwarz J, *et al.* Experience replay for continual learning. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2019. 32.
- 22 Lopez-Paz D, Ranzato MA. Gradient episodic memory for continual learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6470–6479.
- 23 Chaudhry A, Ranzato MA, Rohrbach M, *et al.* Efficient lifelong learning with A-GEM. *Proceedings of the 7th International Conference on Learning Representations*. New Orleans: OpenReview.net, 2019. 1–20.
- 24 Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge: MIT Press, 1986. 318–362.
- 25 Kingma DP, Welling M. Auto-encoding variational Bayes. *Proceedings of the 2nd International Conference on Learning Representations*. Banff: ICLR, 2014. 1–14.
- 26 张建明, 彭锦涛, 贾洪杰, 等. 基于条件变分自编码器的端到端情感语音合成方法. *信号处理*, 2023, 39(4): 678–687. [doi: [10.16798/j.issn.1003-0530.2023.04.009](https://doi.org/10.16798/j.issn.1003-0530.2023.04.009)]
- 27 Serrà J, Suris D, Miron M, *et al.* Overcoming catastrophic forgetting with hard attention to the task. *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018. 4555–4564.

(校对责编: 牛欣悦)