

基于突触可塑性延迟和时间注意力的脉冲神经网络及其语音识别应用

张航铭¹⁾ 白千一¹⁾ 邓智超¹⁾ Alexander Sboev³⁾ 于强^{1,2)}

¹⁾(天津大学智能与计算学部 天津 300072)

²⁾(天津师范大学计算机信息与工程学院 天津 300387)

³⁾(库尔恰托夫研究所国家研究中心 莫斯科 123182 俄罗斯)

摘要 脉冲神经网络(Spiking Neural Networks, SNNs)是一种受大脑启发的计算模型,在处理具有时间和空间维度的数据时具有巨大潜力。研究表明,突触可塑性延迟能够提高SNNs在语音识别等时序任务的性能。然而,时序数据通常具有稀疏性和不均匀性,导致不同时间步中的信息分布存在显著差异。同时,信号中噪声的动态变化进一步加剧了这一问题,使得现有的可塑性延迟方法难以有效捕获关键时间信息,从而限制了模型性能的提升。在这项工作中,我们为可塑性延迟引入了时间注意力机制,通过动态评估各时间步的重要性,进而优化SNNs的性能。具体而言,我们使用扩展卷积来学习延迟和权重。随后,我们在时间维度上利用时间注意力机制动态量化不同时间步的重要性,从而增强模型对关键时间信息的捕获能力。实验结果表明,所提方法在Spiking Heidelberg Digits、Spiking Speech Command和Google Speech Commands数据集上分别取得了96.21%、80.87%和95.60%的精度,相较于传统的可塑性延迟方法有了显著的改进。这一结果验证了我们的方法在处理语音识别任务时的高效性,并为SNNs在处理复杂时序数据中的进一步应用奠定了坚实基础。

关键词 类脑计算;脉冲神经网络;突触可塑性延迟;时间注意力;神经形态计算

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2025.01870

Spiking Neural Networks Based on Synaptic Plasticity Delays and Temporal Attention for Speech Recognition Applications

ZHANG Hang-Ming¹⁾ BAI Qian-Yi¹⁾ DENG Zhi-Chao¹⁾ Alexander Sboev³⁾ YU Qiang^{1,2)}

¹⁾(College of Intelligence and Computing, Tianjin University, Tianjin 300072)

²⁾(College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387)

³⁾(National Research Center Kurchatov Institute, Moscow 123182, Russia)

Abstract As a brain-inspired computational model, Spiking Neural Networks (SNNs) hold great promise for processing data with temporal and spatial dimensions. Research has shown that synaptic plasticity delays can improve the performance of SNNs in temporal tasks such as speech recognition. However, time-series data are usually sparse and inhomogeneous across time steps. This disparity in information distribution, coupled with noise dynamics in the signals, poses challenges for existing plasticity delay methods, which struggle to capture critical temporal information effectively, thereby limiting model performance. In this work, we apply a temporal-wise attention mechanism to plasticity delays, enabling SNNs to evaluate the importance of each

收稿日期:2024-12-07;在线发布日期:2025-05-13。本课题得到国家自然科学基金(No. 92370103, 62176179)、小米公益基金会资助。张航铭,博士研究生,主要研究领域为脉冲神经网络算法、神经编码。E-mail: zhm0755@tju.edu.cn。白千一,硕士研究生,主要研究领域为脉冲神经元学习算法、人工智能。邓智超,硕士研究生,主要研究领域为类脑计算、神经网络。Alexander Sboev,博士,教授,主要研究领域为神经形态计算、忆阻可塑性。于强(通信作者),博士,教授,主要研究领域为类脑智能计算、神经网络、人工智能、理论神经学。E-mail: yuqiang@tju.edu.cn。

time step and optimize performance. Specifically, we use dilated convolution to learn the delay parameters and weights, followed by a temporal attention mechanism that dynamically quantifies the importance of different time steps, thereby enhancing the ability to capture critical temporal information of SNNs. Experimental results show that the proposed method achieves 96.21%, 80.87% and 95.60% accuracy on the Spiking Heidelberg Digits, Spiking Speech Command and Google Speech Commands datasets, respectively, which is a significant improvement over traditional plasticity delay methods. These results highlight the potential of our approach for handling speech recognition tasks and establish a strong foundation for further applications of SNNs in dealing with complex time-series data.

Keywords brain-like computing; spiking neural networks; synaptic plasticity delays; temporal attention; neuromorphic computing

1 引言

脉冲神经网络^[1](Spiking Neural Networks, SNNs)是一种受生物神经系统启发的计算模型,在人工智能领域得到了越来越多的关注。通过模拟神经元之间通过脉冲传递信息的机制,SNNs实现了与人工神经网络^[2-3](Artificial Neural Networks, ANNs)不同的计算特点。其优势不仅体现在对时序数据的高效处理,还包括低功耗和更接近生物学逻辑的建模能力。近年来,SNNs在语音识别^[4-6]等时序任务中表现出色,为神经网络发展提供了新的研究方向。

SNNs的关键特点在于能够精确捕捉和处理输入信号中的时间信息,这使得它们能够区分出数据中的时间动态特征,从而在噪声干扰较大的复杂环境中展现出更强的鲁棒性^[7-8]。在生物神经系统中,神经元之间的信息传递通常伴随着一定的可塑性延迟。这种延迟源于轴突、突触和树突等神经结构,其时间尺度可能长达数十毫秒^[9]。研究表明^[10-11],这种可塑性延迟不仅对神经网络的学习机制至关重要,还为建模生物神经系统的复杂动态计算提供了基础。因此,如何将这种生物启发的延迟机制融入SNNs模型,以优化其在时序任务中的性能并增强其对噪声干扰的适应能力,已成为当前研究的一个关键挑战。这不仅需要针对神经元之间的延迟进行建模,还需解决信噪比变化对时序特征提取的影响,从而进一步提升SNNs在实际应用中的鲁棒性与有效性。

通过攻克这个关键挑战,不仅能够显著提升系统在复杂场景中的整体性能,还能满足诸如智能家居控制和语音助手等实际应用的需求。这些场景要求系统具备快速、精准地解析时序数据并将其转化

为有效指令的能力。更重要的是,通过模拟和利用可塑性延迟,我们能够开发出更符合生物特性的SNNs,从而深入探索大脑处理时间动态信息的机制,推动神经科学与人工智能领域的交叉发展。

近年来,针对延迟机制的研究从多个角度展开,包含了基于生物启发的可塑性学习规则、延迟梯度的计算方法,以及延迟与权重协同优化的策略。例如,文献[12]利用脉冲时间依赖可塑性(Spike Timing Dependent Plasticity, STDP)来学习时间延迟。STDP是一种生物神经网络中广泛存在的突触可塑性机制,它可以根据突触前后神经元的脉冲发放时间差来调整突触权重。通过将STDP机制引入SNNs中,可以让网络自动学习并调整神经元之间的延迟,从而优化网络性能。另一方面,文献[13]提出了一种通过脉冲序列与指数核的卷积来计算损失相对于延迟的梯度的方法。尽管该方法在浅层SNNs中表现良好,但在处理深层SNNs时仍存在一定局限性。深层SNNs的训练通常涉及更多的参数和更复杂的网络结构,这使得训练过程变得更加困难,尤其是在梯度的传播和延迟的精确调整方面。

针对深层SNNs的可塑性延迟,Patiño-Saucedo等人^[14]提出的方法虽然能在神经元修剪前设置固定的延迟突触连接,但这种方法无法同时学习延迟和权重,限制了其在实际应用中的性能。为了克服这一难题,Shrestha等人^[10]提出了一种用于学习突触权重和轴突延迟的新的通用反向传播机制,该机制克服了脉冲发放函数的不可微问题,并使用时间信用分配策略将误差反向传播到前一层。Yu等人^[15]则提出了一种新的联合权重与延迟的可塑性规则,揭示了突触权重和延迟如何相互协作的基本原理,并表明可塑性延迟可以通过跨时间转移信息来增强神经元的选择性和灵活性。最后,Sun等人^[16]验证

了将时间延迟集成到监督学习中的有效性,并提出了一个通过短期记忆调节轴突延迟的模块,从而提高时间特征的表征学习能力。

尽管现有研究在延迟机制的学习与优化上取得了一定进展,但这些方法在处理时序数据时仍然面临显著挑战,主要包括信噪比波动和信息分布不均问题。首先,信噪比波动会导致时间信号的不稳定性,使得模型难以准确捕捉关键时间点。其次,信息分布不均使得某些时间点包含更多关键信息,而其他时间点则相对冗余,传统方法难以有效区分和利用这些关键时间点。这种不均匀性使得SNNs在延迟和权重学习上的策略面临重大挑战,尤其是在处理长时序数据时,模型无法充分利用时间维度上的动态特性,从而限制了性能提升。

为了解决这一问题,我们将时间注意力机制与可塑性延迟设计相结合。该方法的核心思想是通过引入注意力机制,增强SNNs对关键时间节点的关注,从而减少信噪比波动对模型性能的影响。具体而言,时间注意力机制通过计算不同时间步的注意力权重,显式地建模时间维度上的局部特征和依赖关系。其中,局部特征建模使模型能够聚焦于信息量较大的时间窗口,而依赖关系建模则捕捉时间点之间的长程关联。这种设计不仅解决了信息分布不均问题,还显著提升了模型对时序数据中关键信息的利用效率。首先,我们利用可学习扩展卷积^[17]来捕捉和解析延迟参数与权重之间的内在关系。在SNNs中,延迟决定了信息传播的时间特性,而权重则调节信号的强度。通过扩展卷积操作,我们不仅能够有效提取时序数据中的关键特征,还能保留时间维度上的信息,从而优化延迟和权重的协同作用,提高网络对时序特征的建模能力和准确性。然而,直接对稀疏且不均匀的时序数据进行处理仍然是一项具有挑战性的任务。为了降低数据处理的复杂度,我们引入了一种挤压操作,即将每个时间步中的信息聚焦成一个单一的表示点。这一操作有助于将稀疏的时间序列转化为更加紧凑和密集的表示形式,避免了时间轴上的过多数据点造成的处理负担,提升了整体处理效率。接下来,我们进一步引入基于一维卷积的时间注意力机制。在现有的SNNs方法中,时间特征提取通常依赖于固定的时间维度运算或门控机制。例如,STSC-SNN^[18]通过时间响应滤波器和前馈侧向抑制器的双模块架构,实现了任务驱动的特征筛选。其中,时间响应滤波器通过线性计算提取时间特征,而前馈侧向抑制器则通过非

线性门控机制进一步筛选信息。尽管这种方法在时间特征提取上取得了一定效果,但其侧向抑制机制本质上仍依赖于固定的门控策略,难以自适应地聚焦于关键时间节点。相比之下,我们提出的方法通过动态调整不同时间步的注意力权重,使模型能够聚焦于具有高信息熵的关键时间节点,从而实现了更精细化的时间特征表征。

综上所述,我们的主要目标是研究一种能够高效处理稀疏性和不均匀性时序数据的SNNs模型。针对时序数据的特性,我们提出了一种结合可塑性延迟和时间注意力机制的SNNs。实验结果表明,我们的方法在语音识别任务中能够有效提升模型的精度,特别是在噪声较高的语音识别场景中,表现出较传统方法更强的泛化能力和适应性。

本文的其余部分安排如下:第二部分回顾了脉冲神经网络中可塑性延迟与注意力机制的相关研究;第三部分介绍了本文涉及的理论基础;第四部分详细说明了本文提出的方法;第五部分展示了实验设置及其结果的分析比较;最后,第六部分总结了本文的主要结论。

在这项工作中,我们的主要贡献总结如下:

(1)优化了稀疏数据处理与特征提取的方法,通过引入可学习扩展卷积和挤压操作,显著降低了对时序数据的处理复杂度,并将数据转换为紧凑的表示形式。这一优化不仅提升了计算效率,还增强了模型对时间特征的建模能力。

(2)提出了基于突触可塑性延迟和时间注意力机制的创新SNNs模型,该模型能够有效处理稀疏且不均匀的时序数据。通过捕捉关键时间节点并优化延迟与权重的协同作用,模型在复杂时序任务中展现出更高的精度和鲁棒性。

(3)为了验证本文方法的有效性,我们在SHD (Spiking Heidelberg Digits, SHD)、SSC (Spiking Speech Command, SSC) 和 GSC (Google Speech Commands, GSC)数据集上进行了广泛的实验。实验结果表明,该模型在高噪声语音识别任务中表现优异,相较于传统SNNs方法,显著提高了识别精度、泛化能力和适应性。

2 相关工作

2.1 可塑性延迟

在SNNs的研究中,可塑性延迟是一个关键方向,其机制受到生物神经系统中STDP^[19]的启发,同

时也借鉴了经典学习规则^[20]。近年来,研究者们提出了多种优化神经元延迟的策略,这些方法在提升SNNs性能和生物可解释性方面展现出显著效果。首先,Shrestha等人^[10]基于SLAYER学习框架提出了一种延迟优化方法,通过迭代调整神经元轴突的延迟以提升网络性能,这项工作为后续研究奠定了理论和实践基础。随后,Sun等人^[21]进一步改进了该方法,引入可训练的延迟上限,允许网络在训练过程中动态调整每个神经元的延迟范围,从而增强了模型的灵活性和适应性。

在优化策略的多样化方面,研究者们进行了深入探索。Patiño-Saucedo等人^[14]通过对延迟突触的剪枝研究发现,选择性移除部分突触能够减少冗余连接,从而提升网络对数据关键特征的提取能力。Grimaldi等人^[11]则将延迟用于建模多项式逻辑回归,进一步验证了延迟在分类任务中的关键作用。此外,高等人^[22]提出了基于脉冲发放数的间断与连续延迟学习,其中连续延迟学习有效解决了间断方法中延迟调整波动大的问题。

与此同时,一些研究进一步推动了延迟训练方法的创新。例如,Hammouamri等人^[23]提出了一种基于一维卷积的延迟学习方法,其中延迟被表示为可学习的卷积间距。这种设计能够高效编码时间信息,为延迟学习提供了新的技术手段。Wang等人^[13]则通过将脉冲序列转换为连续信号,使得传统梯度下降算法能够直接应用于延迟训练,从而显著提升了训练效率和模型性能。更重要的是,Grappolini等人^[24]的研究表明,仅通过优化延迟而不调整权重,SNNs的性能在某些任务中能够与仅优化权重的模型相媲美。这一发现不仅突出了延迟学习的重要性,也为未来SNNs的设计提供了新的思路。

2.2 注意力机制

在SNNs中,引入注意力机制旨在模拟人脑处理信息的方式,即通过提取关键时间步信息来提高信息处理效率,同时滤除不重要的信息。例如,在处理视觉或听觉信息时,人脑通常仅关注特定区域,而不会对所有输入进行全面理解。这种机制为SNNs的设计提供了新的思路和优化方向。例如,Yao等人^[25]第一次提出在SNNs中引入注意力机制,用以筛选与决策无关的脉冲序列。然而,该方法无法在同一时间步区分不同输入信息,导致空间信息被同等对待。随后,Zhou等人^[26]设计了仅具有空间域自注意力机制的SNNs,但其忽略了不同时间步之间

的信息交互,限制了网络对完整时空特征的建模。为克服上述局限性,Zhu等人^[27]引入时空联合注意力机制,并在神经形态数据集上取得了优异性能。然而,这种方法在设计上破坏了SNNs的异步推理特性。为解决这一问题,Wang等人^[28]提出了一种新的时空自注意力机制,在保留SNNs异步事件驱动特性的同时,能够建模跨时间和空间域的特征依赖性。

后续研究在扩展注意力机制的维度和增强时空感受野方面取得了重要进展,从而显著提升了模型对复杂时空特征的建模能力。例如,Yao等人^[29]提出了多维注意力机制,突破了仅限于一维信息处理的限制,显著提升了模型的适用性。与此同时,Yu等人^[18]受生物突触的启发,设计了一种增强时空感受野的模型,有效提升了对复杂时空依赖关系的建模能力。Xu等人^[30]介绍了DISTA,这是一种具有时空注意力的去噪脉冲Transformer,旨在最大限度地提高脉冲神经元的时空计算能力。Cai等人^[31]受预测性注意力映射实验观察的启发,提出了一种新的空间-通道-时间融合注意力模块,该模块可以引导SNNs利用积累的历史空间-通道信息有效地捕获潜在的目标区域。综上所述,注意力机制的引入为SNNs的时空特征建模提供了强大的工具,推动了SNNs在信息筛选与高效处理方面的提升。

2.3 SNNs的结构设计与优化方法

除了可塑性延迟和注意力机制,SNNs的结构设计和优化方法也是近年来研究的热点方向。在结构设计方面,文献[32]提出基于生物启发的知识蒸馏方法,通过逆向结构模式学习,可以动态优化SNNs中的突触连接以达到最佳状态。此外,Xu等人^[33]提出一种高效的递归脉冲神经网络,以减少原始切片样本的时域信息损失,利用基于脉冲的神经动力学来处理动态时空信息,增强了SNNs对动态时空信息的处理能力,为复杂时序数据的建模提供了新的思路。在模型优化方面,Li等人^[34]提出了一种基于卷积核活动水平的结构化剪枝方法,显著降低了SNNs的计算开销,同时保持了较高的性能。类似地,Shen等人^[35]的工作进一步探索了稀疏选择性激活在持续学习中的应用,为SNNs在动态环境中的适应性提供了技术支持。此外,Jiang等人^[36]提出的自适应SNNs通过平衡兴奋性和抑制性机制,实现了全局-局部学习的协同优化,为SNNs的稳定性和泛化能力提供了新的解决方案。

3 理论基础

3.1 脉冲神经元

目前已有许多模拟生物神经元功能的脉冲神经元模型,如H-H模型^[37]、Izhikevich模型^[38]、LIF模型等。为了更好地平衡计算复杂度和生物合理性,我们使用LIF模型,其膜电位的动态过程可表示为

$$\tau_m \frac{dv_j^{(l)}}{dt} = -(v_j^{(l)}(t) - v_{\text{rest}}) + RI_j^{(l)}(t) \quad (1)$$

膜电位时间常数 τ_m 与膜电阻 R 和膜电容 C 之间满足关系 $\tau_m = R \cdot C$,其中 v_{rest} 表示静息电位。为了简化动力学方程并避免冗余的常数项^[39],常假设 $C=1$ 且 $v_{\text{rest}}=0$,在此设定下,可以通过欧拉方法将(1)式转换成离散时间形式,将神经元发放脉冲后对膜电位的重置过程考虑在内,可表示为

$$v_j^{(l)}[t] = \alpha v_j^{(l)}[t-1] + (1-\alpha) I_j^{(l)}[t] - z_j^{(l)}[t-1] v_{\text{reset}} \quad (2)$$

$$I_j^{(l)}[t] = \sum_i W_{ji}^{(l)} z_i^{(l-1)}[t] \quad (3)$$

其中, $I_j^{(l)}[t]$ 和 $v_j^{(l)}[t]$ 分别表示 t 时刻第 l 层的神经元 j 的输入和膜电位, $W_{ji}^{(l)}$ 表示第 $l-1$ 层的神经元 i 与第 l 层的神经元 j 之间的突触连接的权重,衰减系

数 $\alpha = e^{-\frac{\delta t}{\tau_m}}$,其中 δt 表示离散时间步长。因此第 l 层神经元 j 在 t 时刻的输出 $z_j^{(l)}[t]$ 可以计算为

$$z_j^{(l)}[t] = H(v_j^{(l)}[t] - \vartheta) \quad (4)$$

其中, H 表示Heaviside阶跃函数, ϑ 表示阈值。

3.2 学习算法

在SNNs中,现有的模型训练方法包括STDP^[40]、ANNs转SNNs^[41]、e-prop^[42]、MDGL^[43]以及BPTT^[44]等。STDP根据前后脉冲的时间差来调整突触权重以模拟生物神经突触的可塑性,但难以实现复杂任务的精确控制和优化;ANNs转SNNs方法利用预先训练好的人工神经网络进行转换,但存在精度损失问题;e-prop通过事件驱动方式传播误差梯度,实现了具有生物学可解释性的在线网络学习;MDGL则在e-prop基础上,引入细胞类型特异的神经调节机制,优化了突触信用分配,提高了训练精度。然而,BPTT仍然是目前效果最好的训练方法,其可以直接优化目标函数,具有高效的收敛速度和较高的训练效率。因此,我们选择了BPTT来训练我们的脉冲神经网络,BPTT的主要思想如下:

(2)式已经定义了LIF神经元的膜电位更新方式,为了表达更易理解,可以改写成以下形式:

$$v[t+1] = \underbrace{\alpha v[t]}_{\text{decay}} + \underbrace{(1-\alpha) I[t+1]}_{\text{input}} - \underbrace{z[t] v_{\text{reset}}}_{\text{reset}} \quad (5)$$

在深度学习中,输入的权重通常是可学习的参数,可将(5)式中的 $1-\alpha$ 的影响纳入可学习权重 W 中,并用 W 与输入 $x[t+1]$ 的乘积替换 $I[t+1]$:

$$v[t+1] = \underbrace{\alpha v[t]}_{\text{decay}} + \underbrace{W x[t+1]}_{\text{input}} - \underbrace{z[t] v_{\text{reset}}}_{\text{reset}} \quad (6)$$

其中 $x[t+1]$ 是输入的电压或脉冲,并根据 W 的突触电导进行缩放以生成对神经元的电流输入。反向传播算法使用损失相对于权重的梯度来训练网络,全局的梯度可被定义为权重对历史损失的梯度积累之和,基于链式法可得:

$$\frac{\partial \mathcal{L}}{\partial W} = \sum_t \frac{\partial \mathcal{L}[t]}{\partial W} = \sum_t \frac{\partial \mathcal{L}[t]}{\partial z} \frac{\partial z}{\partial v} \frac{\partial v}{\partial I} \frac{\partial I}{\partial W} \quad (7)$$

由(5)和(6)式可得,(7)式中 $\partial v / \partial I = 1$, $\partial I / \partial W = x$ 。然而 $\partial z / \partial v$ 是一个Dirac Delta函数,其在阈值处趋于无穷大,在其他地方都是0,因此梯度几乎总是为0,无法进行学习。为解决这一问题,我们使用了替代梯度的方法,即在前向计算时保持Heaviside函数不变,在反向传播过程中, $\partial z / \partial v$ 替换为文献[45]中的替代导数 $\Theta'(x)$:

$$\Theta(x) = \frac{1}{\pi} \arctan(\pi x) + \frac{1}{2} \quad (8)$$

$$\frac{\partial z_j^{(l)}[t]}{\partial v_j^{(l)}[t]} = \Theta'(x) = \frac{1}{1 + (\pi x)^2} \quad (9)$$

选择 \arctan 函数作为替代梯度的原因主要有三方面:首先, \arctan 函数具有平滑且连续可微的特性,这有助于优化过程中的稳定性,相较于常见的sigmoid或tanh等函数, \arctan 在输入值极大或极小时提供了更加平缓的梯度变化,从而避免了梯度消失问题。其次, \arctan 的梯度在输入值过大或过小时趋近于零,这一特性有助于防止梯度爆炸。最后,文献[39, 46]中的研究表明, \arctan 替代梯度在SNNs中表现出更好的优化效果。因此,采用 \arctan 作为替代梯度不仅能提升网络训练的稳定性,还能避免常见的梯度问题,优化训练过程。

4 方法

4.1 方法概述

图1展示了我们提出的结合可塑性延迟和时间注意力机制的脉冲神经网络架构。在每个编码

器中, 输入依次经过可塑性延迟模块、时间注意力模块和脉冲神经元层。可塑性延迟模块和时间注意力模块从不同的角度处理信息: 可塑性延迟模块通过调整输入到达的时序, 使时间注意力模块能够更有效地捕捉并区分不同时间步的特征, 从而增强模型对时间维度的感知能力; 时间注意

力模块则通过挤压和激励步骤^[47]计算各时间步信息的重要性, 并在经过可塑性延迟模块调整后的特征表示上, 根据注意力分数对不同时间步信息的重要性进行加权, 从而实现对关键信息的选择性关注, 提升模型对时序数据的辨别与分析能力。

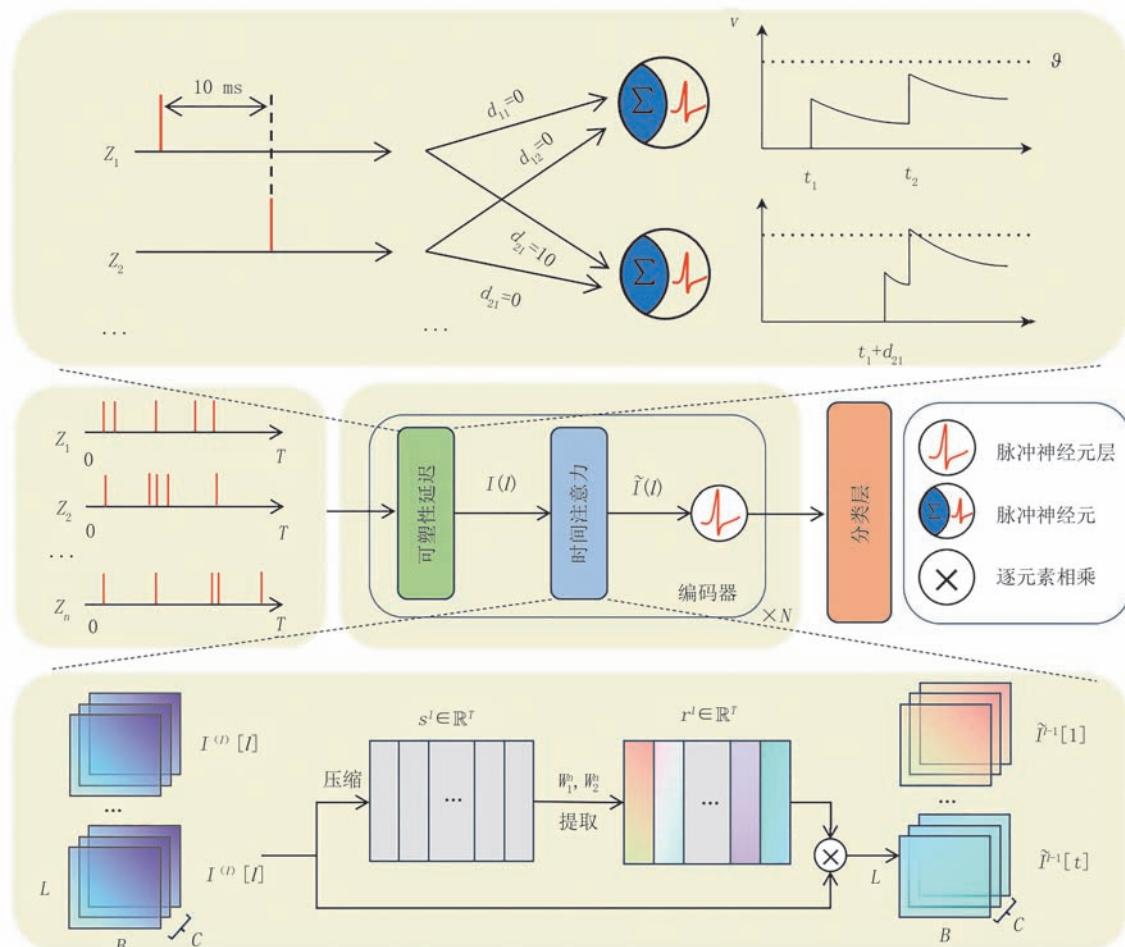


图1 结合可塑性延迟和时间注意力机制的脉冲神经网络框架示意图(输入脉冲信号首先经过多个编码器模块, 每个模块内, 信号通过可学习的延迟机制进行调整。然后, 时间注意力模块根据各时间步信息的重要性对输入信号进行权重分配, 提取关键信息并将优化后的信号传递至脉冲神经元层。最后, 经过多个编码器处理后的数据最终由分类层生成预测结果。)

4.2 可塑性延迟

脉冲从发放到接收之间的时间间隔称为连接延迟, 不同的突触连接之间会有不同的延迟^[9]。我们使用一维时间卷积来模拟连接延迟, 二者之间的数学等价关系已由 Hammouamri 等^[23]证明。如图 2 所示, 在该建模框架下, 每个突触连接都有一个与之对应的一维卷积核。与标准卷积不同, 该卷积核的大小由最大连接延迟决定, 并且仅包含一个非零元素, 该值对应于神经元的突触权重, 其余元素均为零。因此, 该非零元素的数值反映了突触的权重, 而其所

在位置则表示突触的连接延迟。对作用于该突触的输入信号进行卷积, 即可得到经过延迟处理后的每个时间步的输入电流。

带有延迟的前馈 SNNs 模型的参数包括 $\mathbf{W} = (w_{ji}^{(l)}) \in \mathbb{R}$ 和 $\mathbf{D} = (d_{ji}^{(l)}) \in \mathbb{R}^+$, 其中 \mathbf{W} 和 \mathbf{D} 分别表示权重和延迟矩阵, t 时刻第 l 层延迟后的电流为:

$$I_j^{(l)}[t] = \sum_i w_{ji}^{(l)} z_i^{(l-1)}[t - d_{ji}^{(l)}] \quad (10)$$

其中, $w_{ji}^{(l)}$ 和 $d_{ji}^{(l)}$ 分别表示第 $l-1$ 层的神经元 i 与第 l 层的神经元 j 之间的突触连接的权重和延迟, 可将

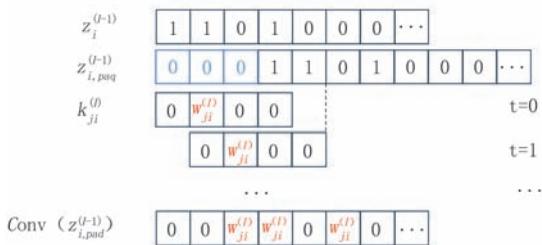


图2 单个突触使用一维时间卷积模拟连接延迟示意图(设 $T_d=3$,对于神经元 j 的第 i 个突触,若其对应的延迟 $d_{ji}^{(l)}=2$,则卷积核 $k_{ji}^{(l)}$ 索引为 $T_d-d_{ji}^{(l)}=1$ 的值为权重 $w_{ji}^{(l)}$,其余位置为0。对 $z_i^{(l-1)}$ 应用大小为 $T_d=3$ 的左填充得到 $z_{i,pad}^{(l-1)}$ 。蓝色框表示填充部分。再对 $z_{i,pad}^{(l-1)}$ 进行卷积操作(卷积的步长为1,与时间步长无关),得到该突触延迟后的电流,将所有突触的电流求和, $I_j^{(l)}=\sum_{i=1}^n \text{Conv1D}(z_{i,pad}^{(l-1)})$,即可得到神经元 j 的电流 $I_j^{(l)}$ 。)

其建模为一维时间卷积的卷积核 $k_{ji}^{(l)}$:

$$k_{ji}^{(l)}[n]=\begin{cases} w_{ji}^{(l)}, & \text{if } n=T_d-d_{ji}^{(l)} \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

其中, $n \in [[0, \dots, T_d]]$, T_d 表示最大连接延迟数。为使 $I[0]$ 对应于 $t=0$,我们在 z 上使用了值为0,大小为 T_d 的左填充,记为 $z_{i,pad}$,因此,(10)式可重写为卷积结果之和:

$$I_j^{(l)}=\sum_i \text{Conv1D}(z_{i,pad}^{(l-1)}) \quad (12)$$

突触的权重及其位置通过 Khalfaoui-Hassani 等^[48]提出的可学习间隔扩展卷积方法进行学习:在训练过程中,由于一维卷积核的稀疏性,为了能够计算损失 \mathcal{L} 对延迟 $d_{ji}^{(l)}$ 的导数 $\frac{d\mathcal{L}}{dd_{ji}^{(l)}}$,使用均值为 $T_d-d_{ji}^{(l)}$ 、标准差为 $\sigma_{ji}^{(l)} \in R^*$ 的高斯核^[48]来初始化卷积核元素,其中 c 是归一化项,取 $\epsilon=1e-7$ 以避免除以零。在这种初始化方式下,高斯分布曲线下的面积对应于突触的权重 $w_{ji}^{(l)}$ 。

$$k_{ji}^{(l)}[n]=\frac{w_{ji}^{(l)}}{c} \exp\left(-\frac{1}{2}\left(\frac{n-T_d+d_{ji}^{(l)}+2}{\sigma_{ji}^{(l)}}\right)^2\right) \quad (13)$$

$$c=\epsilon+\sum_{n=0}^{T_d} \exp\left(-\frac{1}{2}\left(\frac{n-T_d+d_{ji}^{(l)}+2}{\sigma_{ji}^{(l)}}\right)^2\right) \quad (14)$$

参照文献[23]中的训练和测试方案,在训练过程中,每个批次结束后对 $d_{ji}^{(l)}$ 进行裁剪操作,以确保其值保持在0与 T_d 之间。在训练开始时设定 $\sigma=\left\lfloor\frac{T_d+1}{2}\right\rfloor$,并使用(13)式初始化卷积核。在验证集或测试集上进行评估时,将卷积核转换为(11)式形

式的离散核。

通过一维时间卷积对突触延迟进行显式建模并学习延迟及权重,核心思想是将时间维度上的信号传递延迟转化为卷积核空间位置的可学习参数, $w_{ji}^{(l)}$ 和 $d_{ji}^{(l)}$ 分别直接决定从神经元 i 到神经元 j 的突触权重和延迟时间步,从而实现了时序信息与突触权重的联合编码。(13)式的初始化方式克服了离散延迟参数 $d_{ji}^{(l)}$ 的不可微性,使得延迟与权重可通过标准反向传播优化。当网络训练完成后,不同突触在卷积核上的非零元素位置彼此之间的差异,正体现了网络在全局范围内对时序信息的挖掘与调整。

4.3 时间注意力机制

传统 Transformer 的自注意力机制^[49]主要通过查询、键和值的点积计算来捕捉输入序列中元素之间的全局依赖关系,其核心优势在于对空间维度的建模能力。然而,这种机制在处理时序数据时存在明显局限性:时间维度的信息仅通过位置编码引入,缺乏对时间动态特性的显式建模,难以捕捉时间序列中关键时间点的重要性。此外,传统自注意力机制的计算复杂度为 $O(n^2)$,在处理长序列时计算开销较大。针对这些问题,我们引入了一种新的时间注意力机制,专门用于 SNNs 中的时序数据处理。该方法通过引入时间注意力权重,直接计算时间序列中不同时间点的重要性,从而显式建模时间维度上的动态特性。此外,时间注意力机制通过挤压和激励步骤降低了计算复杂度,避免了传统方法中全局注意力矩阵的高内存占用问题。

在挤压阶段,通过计算每个时间步的特征向量的平均值,将每个时间步中的信息聚焦成一个单一的表示点 $s^{(l)}[t]$,以降低 SNNs 对时序数据的处理复杂度,公式化为

$$s^{(l)}[t]=\frac{1}{C \times L \times B} \sum_{k=1}^C \sum_{i=1}^L \sum_{j=1}^B \mathbf{I}^{(l)}[\mathbf{t}](k, i, j) \quad (15)$$

其中, $\mathbf{I}^{(l)}[\mathbf{t}] \in \mathbb{R}^{C \times L \times B}$ 是第 l 层在时刻 t 的空间输入张量, C 是通道数, $L \times B$ 代表空间分辨率。

在激励阶段,通过两层全连接网络对不同时间步信息的重要性进行建模,从而生成每个时间步的注意力分数 $\mathbf{r}^{(l)}$ 。最后将 $\mathbf{r}^{(l)}$ 与输入的特征进行逐元素相乘,实现对特征的加权调整。相关计算公式如下:

$$\mathbf{r}^{(l)}=S(\mathbf{W}_2^{(l)} \text{ReLU}(\mathbf{W}_1^{(l)} \mathbf{s}^{(l)})) \quad (16)$$

其中, S 代表 Sigmoid 激活函数, $\mathbf{W}_1^{(l)} \in \mathbb{R}^{T \times T}$ 和

$\mathbf{W}_2^{(l)} \in \mathbb{R}^{T \times \frac{T}{y}}$ 是两个可由 y 控制大小的参数矩阵。当 y 增大时, $\mathbf{W}_1^{(l)}$ 和 $\mathbf{W}_2^{(l)}$ 的维度变小, 模型的可训练参数量随之减少。最后, 将 $\mathbf{r}^{(l)}$ 作为得分向量, t 时刻最终的输入电流可表示为

$$\tilde{\mathbf{I}}^{(l)}[\mathbf{t}] = \mathbf{r}^{(l)}[\mathbf{t}] \cdot \mathbf{I}^{(l)}[\mathbf{t}] \quad (17)$$

在注意力分数的计算过程中, 首先将每个时间步的空间特征聚合成向量 $\mathbf{s}^{(l)}$, 然后利用参数矩阵 $\mathbf{W}_1^{(l)}$ 和 $\mathbf{W}_2^{(l)}$ 对整段时间上的信息进行非线性映射, 使得模型能够从整段时间序列的全局特征中学习到反映不同时间步重要性的权重向量 $\mathbf{r}^{(l)}$ 。这意味着每个时间步的注意力得分不只是依赖其自身特征, 还会综合考虑其他时间步的信息, 通过这种方式, 模型能够对重要时间步给予更高的关注度。

5 实验与分析

5.1 数据集

本文采用三个SNNs领域常用的语音识别基准数据集进行实验: 面向连续语音识别的原始音频数据集 Google Speech Commands v0.02(GSC)^[50], 以及经过人工耳蜗脉冲编码的SHD和SSC数据集^[51]。GSC数据集包含十万条不同的录音, 由35类常用语音命令(如“yes”、“no”、“up”等)组成。该数据集作为智能语音交互领域的重要基准, 包含由数千名志愿者通过移动设备采集的短语音命令, 采样频率为16 kHz, 每条语音长度严格控制在1秒以内, 为脉冲神经网络的时序特征提取提供了理想素材。SHD数据集由一系列脉冲序列数据构成, 包含数字0~9的英语或德语口述发音, 采样频率为16 kHz, 共计20个类别。数据集录音采集于12个志愿者, 其中2个志愿者的录音仅在测试集中出现, 共包含10 000条不同的录音。SHD和SSC两个数据集的构造过程均是借助人工耳蜗将原始音频数据通过相位编码(Phase Coding)转化为脉冲数据, 原始音频数据集分别是Heidelberg Dataset(HD)和GSC。数据集每段录音的原始声波数据都已经转换为700个输入通道上的脉冲序列数据。通过使用这三个数据集, 我们能够全面评估所提出的基于突触可塑性延迟和时间注意力机制的SNNs在语音识别任务中的有效性和优势。数据集的详细统计信息如表1所示。

5.2 实验设置

本文使用开源工具 SpikingJelly^[52]构建语音识别网络。我们使用多个编码器模块相连接构建网络

表1 数据集统计信息

数据集	SHD	SSC/GSC
训练样本数	8332	65 466
测试样本数	2088	9981(验证)/20 382(测试)
类别	20	35
通道数	700	700

模型, 每个编码器由多个子模块组成。我们使用可学习间隔扩展卷积来构造突触前神经元和突触后神经元之间的每个突触连接, 即突触可塑性延迟模块。突触输出经过批归一化后, 输入一维时间注意力模块, 以增强网络对关键时间节点的关注, 最终进入脉冲神经元层进行脉冲计算。每个编码器的末尾添加了dropout模块, 以增强模型的泛化能力。

训练过程中, 我们采用 Adam 优化器^[53]、OneCycle 周期性学习策略^[54]调整学习率。初始学习率设定为 0.001。模型的训练配置针对不同数据集进行调整: GSC 采用 100 个 epoch, SHD 与 SSC 分别设置 150 和 60 个 epoch, batch size 统一设置为 256。表 2 列出了实验中使用的编码器数量以及超参数, 其中 N 表示编码器数量, n 表示脉冲神经元层的大小, τ_m 指神经元的膜电位时间常数, y 用于控制模型可训练参数量。我们使用 Hammouamri^[23]等人的超参数选取策略, 为防止过拟合风险采用较浅的网络层数和每层较小的神经元数量。结合生物合理性, 脉冲神经元的时间常数选择常见范围值 10~15 ms。考虑到 SNNs 的脉冲稀疏性自带正则化效果, 搭建网络时设置了较低的 Dropout Rate。经过对数据集语音时长的统计分析, 采用语音平均时长作为各个数据集的最大延迟数(T_d)。

表2 网络结构和超参数

数据集	SHD	SSC/GSC
N	2	3
n	256	512
τ_m (ms)	10	15
T_d (ms)	250	300
Dropout Rate	0.4	0.25
y	5	5

5.3 结果与分析

语音识别结果如表 3 所示, 实验数据通过对三次独立运行结果取算术平均值获得, 并标注标准差 (Mean \pm SD)。我们的方法在 SHD 数据集上取得了 96.21% 的分类精度, 显著优于其他方法。与仅采用时间注意力机制但未引入可塑性延迟的

表3 在SHD、SSC和GSC数据集上的分类准确率

数据集	模型	注意力	延迟	循环	准确率/%
SHD	EventProp-GeNN ^[55]	×	×	✓	84.80±1.50
	Cuba-LIF ^[56]	×	×	✓	87.80±1.10
	Adaptive SRNN ^[57]	×	×	✓	90.40
	SNN+Delays ^[14]	×	✓	×	90.43
	TA-SNN ^[25]	✓	×	×	91.08
	STSC-SNN ^[18]	✓	×	×	92.36
	Adaptive Delays ^[16]	×	✓	×	92.45
	DL128-SNN-Dloss ^[5]	×	✓	×	92.56
	Dense Conv Delays ^[23]	×	✓	×	93.44
	RadLIF ^[58]	×	×	✓	94.62
SSC	DCLS-Delays(2L-1KC) ^[23]	×	✓	×	95.07±0.24
	Ours	✓	✓	×	96.21±0.12
	Recurrent SNN ^[51]	×	×	✓	59.90±1.10
	Heter. RSNN ^[7]	×	×	✓	57.30
	SNN-CNN ^[59]	×	✓	×	72.03
	Adaptive SRNN ^[57]	×	×	✓	74.20
	SpikeGRU ^[56]	×	×	✓	77.00±0.40
	RadLIF ^[58]	×	×	✓	77.40
	Dense Conv Delays 3L ^[23]	×	✓	×	78.44
	DCLS-Delays(3L-2KC) ^[23]	×	✓	×	80.69±0.21
GSC	Ours	✓	✓	×	80.87±0.06
	MSAT ^[60]	×	×	×	87.33
	Dense Conv Delays ^[23]	×	✓	×	93.19
	RadLIF ^[58]	×	×	✓	94.51
	DCLS-Delays(3L-2KC) ^[23]	×	✓	×	95.35±0.04
	Ours	✓	✓	×	95.60±0.04

STSC-SNN 方法^[18]相比,我们的方法提升了 3.85%,这表明了可塑性延迟在语音识别任务中的突出作用。可塑性延迟通过调整突触的响应时间,使模型更好地适应复杂的时间序列,从而提高识别的准确性。相比之下,尽管 DCLS-Delays (2L-1KC)方法^[23]同样采用了可塑性延迟,但我们的方法仍然有着 1.14% 的性能优势,这一实验结果证明了时间注意力机制能够有效捕捉到输入序列中的重要时间步。值得注意的是,我们的方法在未引入额外循环结构的情况下,与采用循环结构的 RadLIF 方法^[58]相比,性能提升了 1.59%。循环连接结构通常有助于捕捉长时间依赖信息,但也增加了模型的复杂度和参数量,而我们的模型仅依赖 LIF 神经元的内在循环连接,仍然在分类精度上表现出色。

在SSC数据集上,我们的方法实现了80.87%的分类精度。相比仅使用循环结构但未引入可塑性延迟的RadLIF方法^[58]提高了3.47%。本文方法在SHD和SSC数据集上相较于RadLIF的领先优势,

说明可塑性延迟机制与时间注意力机制的结合同样能够有效捕捉并利用长时间依赖信息。此外,与使用可塑性延迟的DCLS-Delays(3L-2KC)模型^[23]相比,我们的方法在分类准确率上提高了0.18%。当应用于原始音频数据时,本方法在GSC数据集上取得95.60%的分类精度,不仅以8.27%的显著优势超越MSAT模型^[60],相较DCLS-Delays(3L-2KC)模型^[23]也保持0.21%的领先幅度。这些跨数据集的性能提升展现出模型对不同编码方式音频信号更强的泛化能力,进一步验证了可塑性延迟和时间注意力机制在处理语音识别任务中的有效性,突显了本文方法在复杂时序任务中的潜力。

图3展示了训练中使用的语音信号样本,以及神经网络生成的注意力图和训练前后可塑性延迟核心的变化情况,进一步揭示了我们方法的工作机制。图3(a)展示了语音信号在时间轴上的分布情况,并标示了与实际语音信号“ZERO”发音位置的对应关系。图3(b)展示了经过训练后的注意力机制对输入信号时间节点的关注情况。图中时间步颜色深浅反映了模型对该时刻输入信息的关注强度。可以观察到,模型对语音信号的关键发音时刻给予了更多的关注,从而在处理语音识别等复杂时序任务时更具优势。图3(c)和(d)展示了神经网络在训练前后突触延迟核心的变化情况。图中的核心位置表示对应突触的延迟时间步,而颜色深浅则反映了学习权重的大小。该结果展现了可塑性延迟机制学习信号时间特征的过程。通过有效结合时间注意力机制与可塑性延迟,网络能够更好地捕捉到输入信号的时间特征,更精准地关注重要的时间节点,从而在处理语音识别任务时获得进一步的性能提升。

5.4 消融实验

为了进一步验证可塑性延迟和时间注意力机制的有效性,我们在SHD数据集上设计了一系列消融实验。

5.4.1 协同性探讨实验

首先我们针对可塑性延迟和时间注意力机制的协同性设计了实验。在实验中,我们采用没有可塑性延迟和时间注意力机制的前馈SNNs作为基准模型(Baseline)。接着,我们逐步将可塑性延迟(基于扩展卷积实现)和时间注意力机制引入基准模型,以观察它们对模型性能的影响。

实验结果如表4所示,表中“w\o D”表示没有使用可塑性延迟机制的方法,“w\o A”表示没有使用时间注意力机制的方法。实验数据显示,拥有可塑

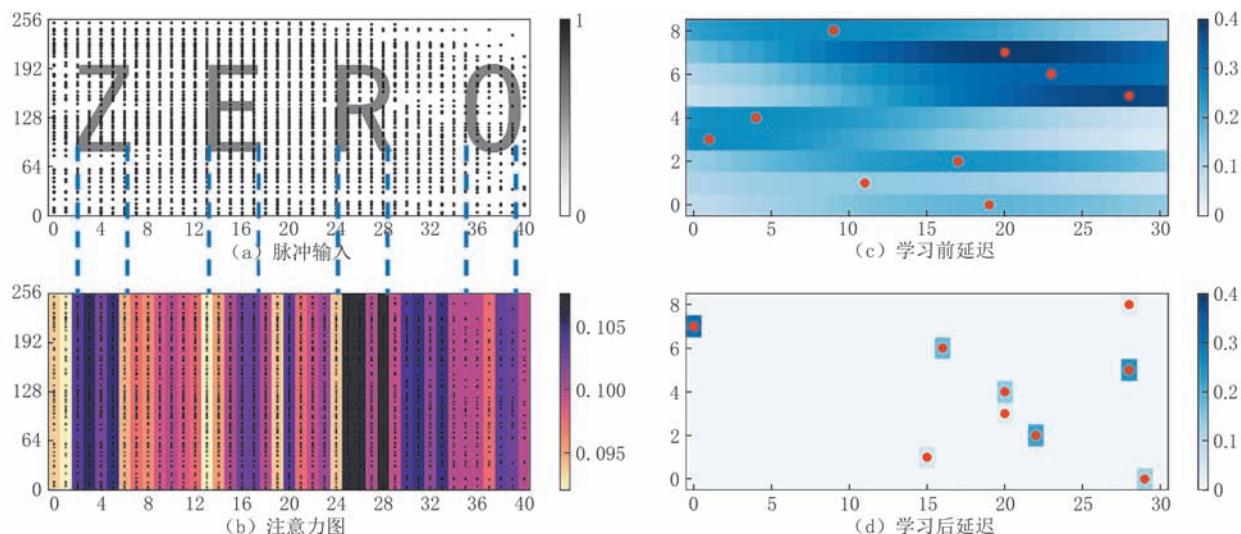


图3 神经网络时间注意力分布图及训练前后延迟核心变化示意图((a)图表示原始脉冲输入。横轴表示时间步长,纵轴表示神经元。(b)图表示学习后的时间维度注意力分布。(c)图和(d)图分别展示了随机选择的8个突触在学习前后延迟核心变化示例。横轴表示时间步长,纵轴表示突触。延迟核心由红点标出。颜色深浅表示权重大小。)

性延迟的模型在分类精度上均显著高于无延迟机制的模型。这一结果与我们的理论预期相符,表明引入可塑性延迟的模型能够成功捕捉和解析数据中延迟参数与可学习权重之间的深层关系,有效建模时序数据的动态变化,实现更高的分类准确率。

表4 模块协同性探讨实验

数据集	模型	延迟	注意力	准确率/%
SHD	Baseline	×	×	65.28
	w\o D	×	✓	76.33
	w\o A	✓	×	95.29
	Ours	✓	✓	96.21

此外,仅拥有时间注意力机制的模型在准确率上也比基础模型高出了11.05%。这一性能提升充分验证时间注意力机制能够聚焦最关键的时间步。最后,我们将可塑性延迟和时间注意力机制同时集成到模型中,模型的性能取得了最高的分类精度。这一结果验证了这两种机制在脉冲神经网络中的有效性,而且展示了它们在语音识别任务中的协同作用。具体分析这一结果,我们发现,可塑性延迟机制通过对突触延迟的自适应调整,使模型能够精准地对齐输入信号的时间步长,从而有效应对语音信号中存在的时间不对齐问题。通过在训练过程中不断优化突触延迟,模型能够灵活地适应不同语音信号的时间特性。与此同时,时间注意力机制通过动态调整模型在不同时间步信息上的关注度,使得模型能够重点捕捉对语音识别至关重要的时序信息。

5.4.2 模块结构探讨实验

在本次实验中,我们构建了AFDM(Attention-First Delay Module, AFDM)、AMDM(Attention-Middle Delay Module, AMDM)、ALDM(Attention-Last Delay Module, ALDM)和PDAM(Parallel Delay-Attention Module, PDAM)四个实验模型,系统性地探索可塑性延迟与时间注意力机制的空间顺序对模型准确性的影响。图4展示了各个实验模型的模块空间顺序,其中基准模型(Baseline)^[23]仅包含可塑性延迟结构,没有时间注意力机制的引入。AFDM将时间注意力机制连接到基准模型之前,旨在评估注意力机制是否能在延迟信息被传递之前优化信息选择。AMDM将时间注意力机制置于可塑性延迟和脉冲神经元之间,评估其在捕捉和解析延迟信息中关键时间步的能力。ALDM将时间注意力机制放在基准模型之后使用,以分析神经元处理脉冲信息后再引入注意力机制是否能提升信息选择性和模型性能。PDAM将可塑性延迟结构和注意力机制结构并行连接,两者输出合并后连接到脉冲神经元层,旨在探讨并行结构能否提升两种机制协同处理信息的效率与精度。

实验结果如表5所示,AMDM模型在SHD和SSC数据集上分别获得了96.21%和80.87%的最高准确率。这些数据表明,AFDM将注意力机制放在延迟之前会影响延迟信息的完整性,而ALDM对神经元输出信息进行处理,导致注意力机制无法关注到关键的输入时间步信息。PDAM虽然并行连

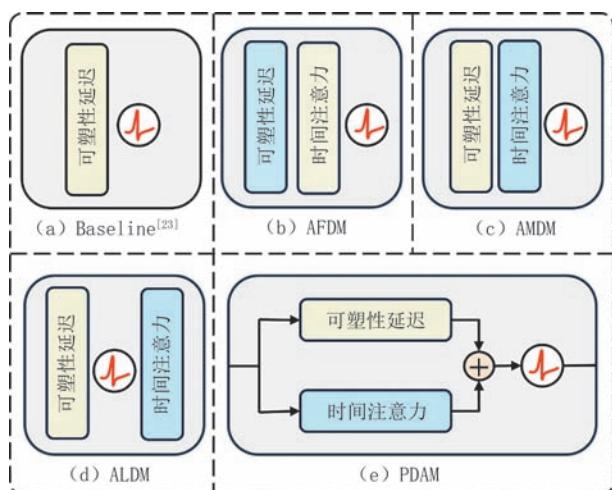


图4 模块结构示意图(图(a)展示了基准DCLS-Delays模型[23]。在此基础上,我们构建了多个改进模型。具体而言,图(b)、(c)和(d)采用串行结构,将时间注意力机制分别放置于基准模型的前(AFDM)、中(AMDM)、后(ALDM)位置。图(e)采用并行连接结构将可塑性延迟和时间注意力的输出合并。 \oplus 表示矩阵逐元素相加。)

接了两者,但未能考虑信息处理的时序性,导致性能损失。AMDM避免了注意力机制过早或过晚介入对信息处理的潜在影响。实验结果表明将时间注意力机制作为可塑性延迟和脉冲神经元的中间结构,能够充分捕捉和解析数据集中的关键时间步信息,提高网络对复杂语音信号的建模能力和分类准确性。因此,我们选择将AMDM模型作为编码器的基础模块来构建神经网络并用于其他实验探索。

综上所述,本实验深刻揭示了在网络架构中可塑性延迟与时间注意力机制的空间顺序对于提升模型性能的重要性。特别是在处理复杂数据集时,通过

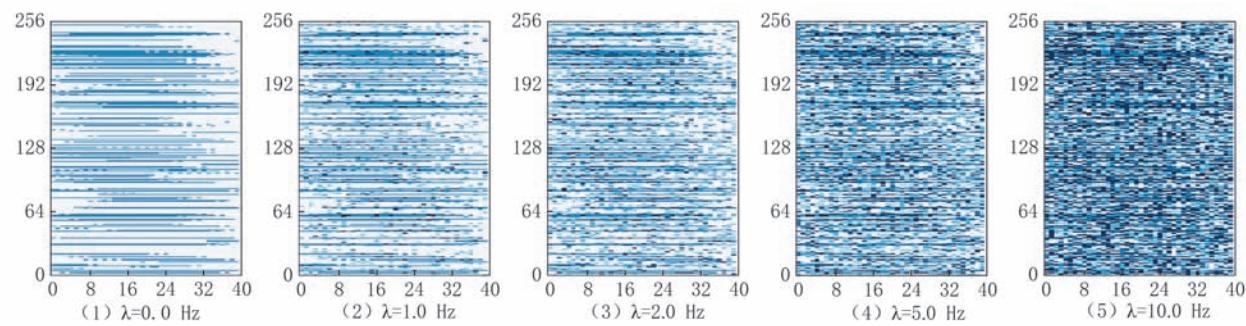


图5 噪声数据对比图(比较了无噪声($\lambda=0.0$ Hz)与添加不同程度噪声后的输入数据表示。横轴表示时间步长,纵轴表示神经元。)

鲁棒性测试实验结果如表6所示。在SHD数据集上,基准模型在无噪声条件($\lambda=0.0$ Hz)下的分类准确率为95.07%,而随着噪声水平的增加,模

表5 模块协同性探讨实验

数据集	模型	相对位置	准确率/%
SHD	Baseline ^[23]	-	95.07
	AFDM	First	94.55
	AMDM	Middle	96.21
	ALDM	Last	94.25
SSC	PDAM	Parallel	95.31
	Baseline ^[23]	-	80.69
	AFDM	First	80.15
	AMDM	Middle	80.87
SSC	ALDM	Last	80.32
	PDAM	Parallel	80.32

精细调控模块间的空间顺序可以有效优化信息处理流程,提高模型的准确性。这为未来设计更加高效、适应性更强的神经网络架构提供了有力的实践依据。

5.5 鲁棒性测试

SNNs能够精准捕捉和处理时间信息,因此在面对存在噪声干扰的数据时表现出较强的鲁棒性,为此,我们设计了鲁棒性对比实验,证明本文方法能够增加SNNs的抗干扰性质。

在本次鲁棒性实验中,我们系统地评估了模型在SHD数据集上面对不同强度噪声的鲁棒性表现。为了模拟突触传输噪声或传感器信号失真等现实场景中的干扰,本研究采用文献[44, 61-62]提出的方法,通过在输入脉冲序列中叠加符合泊松分布的随机噪声来模拟不同强度的环境扰动。如图5所示,其中 $\lambda=0.0$ Hz表示未加入任何噪声。通过分析模型在不同噪声水平下的准确率变化,我们得以评估模型的抗噪能力。我们使用文献[23]中的DCLS-Delays(2L-1KC)模型作为基准对比模型(Baseline)。

型准确率逐步下降。当噪声强度达到 $\lambda=5.0$ Hz时,基准模型的准确率下降了23.24%。这一结果表明,基准模型具有一定的抗噪能力。相比之下,我

们的模型在无噪声条件下的准确率为 96.21%。虽然在加入噪声后准确率也有所下降,但在所有噪声条件下,我们的模型准确率始终优于基准模型,显示出更强的鲁棒性。

表 6 鲁棒性测试实验结果

数据集	λ (Hz)	Baseline/Ours/%
SHD	0.0	95.07/ 96.21(+1.14)
	1.0	89.63/ 92.19(+2.56)
	2.0	71.83/ 83.74(+11.91)
	5.0	65.31/ 82.23(+17.92)
	10.0	9.16/ 12.33(+3.17)

6 总结与展望

本文提出了一种结合可塑性延迟和时间注意力机制的创新性SNNs模型,旨在应对时序数据稀疏性和不均匀性对模型精度的挑战。实验结果表明,该方法在语音识别任务中表现出显著优势。在SHD、SSC和GSC数据集上,模型的分类准确率分别达到96.21%、80.87%和95.60%,显著优于现有方法,充分验证了其有效性。消融实验进一步揭示了可塑性延迟与时间注意力机制的协同作用相较于单一机制的显著优势,结合后模型准确率分别较单一机制提升了19.88%和0.92%。此外,我们系统性地探讨了两种机制的不同组合方式,实验结果显示,将时间注意力机制置于可塑性延迟和脉冲神经元层之间能够更有效地捕获关键特征,为未来网络架构的设计和优化提供了重要的实验依据。

进一步的实验结果表明,该模型在面对噪声数据时表现出极高的稳定性和鲁棒性,具体而言,在添加符合参数为2Hz的泊松分布噪声后仍能保持83.74%的准确率,远优于其他方法。这一特性彰显了模型在实际应用场景中的可靠性和适用性。然而,本研究的局限性在于实验验证仅覆盖了三个数据集,难以全面评估模型的泛化能力。为了应对更复杂的数据集,我们计划进一步优化时间注意力机制。具体来说,可以通过引入更高效的计算方法,如稀疏化技术和低秩近似,来降低计算开销并提高效率。此外,采用局部时间注意力机制,可以减少全局计算的复杂度,确保模型在大规模数据集上的可扩展性。在硬件实现方面,SNNs的低能耗特性使其在神经形态计算领域具有显著优势。未来,我们还将探索将时间注意力机制与神经形态硬件相结合,以充分发挥其能效优

势。具体而言,我们计划研究硬件友好的算法设计,例如基于事件驱动的注意力机制,以减少计算开销并适配神经形态硬件的并行计算特性。同时,我们将在实际的神经形态硬件平台上验证模型的性能,评估其在能效、延迟和准确性方面的表现。

作者贡献声明 张航铭、白千一与邓智超在本研究中贡献相同,为本文的共同第一作者。

参 考 文 献

- [1] Zhang Tie-Lin, Xu-Bo. Research advances and perspectives on spiking neural networks. Chinese Journal of Computers, 2021, 44(9):1767-1785 (in Chinese)
(张铁林, 徐波. 脉冲神经网络研究现状及展望. 计算机学报, 2021, 44(9): 1767-1785)
- [2] WU J, GAUR Y, CHEN Z, et al. On decoder-only architecture for speech-to-text and large language model integration//Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop. Taipei, China, 2023: 1-8
- [3] RADFORD A, KIM J W, XU T, et al. Robust speech recognition via large-scale weak supervision//Proceedings of the International Conference on Machine Learning. Honolulu, USA, 2023: 28492-28518
- [4] WU J, YILMAZ E, ZHANG M, et al. Deep spiking neural networks for large vocabulary automatic speech recognition. Frontiers in Neuroscience, 2020, 14: 513257
- [5] SUN P, CHUA Y, DEVOS P, et al. Learnable axonal delay in spiking neural networks improves spoken word recognition. Frontiers in Neuroscience, 2023, 17: 1275944
- [6] YU C, DU Y, CHEN M, et al. MAP-SNN: Mapping spike activities with multiplicity, adaptability, and plasticity into biologically plausible spiking neural networks. Frontiers in Neuroscience, 2022, 16: 945037
- [7] PEREZ-NIEVES N, LEUNG V C, DRAGOTTI P L, et al. Neural heterogeneity promotes robust learning. Nature Communications, 2021, 12(1): 5791
- [8] CHENG X, HAO Y, XU J, et al. LISNN: Improving spiking neural networks with lateral interactions for robust object recognition//Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama, Japan, 2020: 1519-1525
- [9] IZHKEVICH E M. Polychronization: Computation with spikes. Neural Computation, 2006, 18(2): 245-282.
- [10] SHRESTHA S B, ORCHARD G. Slayer: Spike layer error reassignment in time. Advances in Neural Information Processing Systems, 2018, 31
- [11] GRIMALDI A, PERRINET L U. Learning heterogeneous delays in a layer of spiking neurons for fast motion detection. Biological Cybernetics, 2023, 117: 373-387
- [12] HAZAN H, CABY S, EARL C, et al. Memory via temporal delays in weightless spiking neural network. arXiv preprint

- arXiv:2202.07132, 2022
- [13] WANG X, LIN X, DANG X. A delay learning algorithm based on spike train kernels for spiking neurons. *Frontiers in Neuroscience*, 2019, 13: 429572
- [14] PATIÑO-SAUCEDO A, YOUSEFZADEH A, TANG G, et al. Empirical study on the efficiency of spiking neural networks with axonal delays, and algorithm-hardware benchmarking// *Proceedings of the IEEE International Symposium on Circuits and Systems. Virtual*, 2023: 1-5
- [15] YU Q, GAO J, WEI J, et al. Improving multispike learning with plastic synaptic delays. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 34(12): 10254-10265
- [16] SUN P, ZHU L, BOTTELOOREN D. Axonal delay as a short-term memory for feed forward deep spiking neural networks// *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes, Greece*, 2022: 8932-8936
- [17] YU F. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015
- [18] YU C, GU Z, LI D, et al. Stsc-snn: Spatio-temporal synaptic connection with temporal convolution and attention for spiking neural networks. *Frontiers in Neuroscience*, 2022, 16: 1079357
- [19] WANG R, COHEN G, STIEFEL K M, et al. An FPGA implementation of a polychronous spiking neural network with delay adaptation. *Frontiers in Neuroscience*, 2013, 7: 14
- [20] ZHANG M, WU J, BELATRECHE A, et al. Supervised learning in spiking neural networks with synaptic delay-weight plasticity. *Neurocomputing*, 2020, 409: 103-118
- [21] SUN P, EQLIMI E, CHUA Y, et al. Adaptive axonal delays in feedforward spiking neural networks for accurate spoken word recognition// *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes, Greece*, 2023: 1-5
- [22] Gao Jia-Lu, Yu Qiang, Tang Hua-Jin, et al. Research on efficient multi-spike learning algorithms joint with synaptic weight and delay plasticity. *Chinese Journal of Computers*, 2022, 45(10):2065-2079 (in Chinese)
(高嘉潞, 于强, 唐华锦 等. 联合突触权重和延迟可塑性的高效多脉冲学习算法研究. *计算机学报*, 2022, 45(10): 2065-2079)
- [23] HAMMOUARI I, KHALFAOUI-HASSANI I, MASQUELIER T. Learning delays in spiking neural networks using dilated convolutions with learnable spacings. *arXiv preprint arXiv:2306.17670*, 2023
- [24] GRAPPOLINI E W, SUBRAMONEY A. Beyond weights: Deep learning in spiking neural networks with pure synaptic-delay training// *Proceedings of the 2023 International Conference on Neuromorphic Systems. Santa Fe, USA*, 2023: 1-4
- [25] YAO M, GAO H, ZHAO G, et al. Temporal-wise attention spiking neural networks for event streams classification// *Proceedings of the IEEE/CVF International Conference on Computer Vision. Online*, 2021: 10221-10230
- [26] ZHOU Z, ZHU Y, HE C, et al. Spikformer: When spiking neural network meets transformer// *Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda*, 2023: 1-17
- [27] ZHU R J, ZHANG M, ZHAO Q, et al. Tcja-snn: Temporal-channel joint attention for spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(3): 5112-5125
- [28] WANG Y, SHI K, LU C, et al. Spatial-temporal self-attention for asynchronous spiking neural networks// *Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Macao, China*, 2023: 3085-3093
- [29] YAO M, ZHAO G, ZHANG H, et al. Attention spiking neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(8): 9393-9410
- [30] XU B, GENG H, YIN Y, et al. DISTA: Denoising Spiking transformer with intrinsic plasticity and spatiotemporal attention. *arXiv preprint arXiv:2311.09376*, 2023
- [31] CAI W, SUN H, LIU R, et al. A spatial-channel-temporal-fused attention for spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(10): 14315-14329
- [32] XU Q, LI Y, FANG X, et al. Reversing structural pattern learning with biologically inspired knowledge distillation for spiking neural networks// *Proceedings of the 32nd ACM International Conference on Multimedia. New York, USA*, 2024: 3431-3439
- [33] XU Q, FANG X, LI Y, et al. RSNN: Recurrent spiking neural networks for dynamic spatial-temporal information processing// *Proceedings of the 32nd ACM International Conference on Multimedia. New York, USA*, 2024: 10602-10610
- [34] LI Y, XU Q, SHEN J, et al. Towards efficient deep spiking neural networks construction with spiking activity based pruning. *arXiv preprint arXiv:2406.01072*, 2024
- [35] SHEN J, NI W, XU Q, et al. Efficient spiking neural networks with sparse selective activation for continual learning// *Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada*, 2024: 611-619
- [36] JIANG T, XU Q, RAN X, et al. Adaptive deep spiking neural network with global-local learning via balanced excitatory and inhibitory mechanism// *Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria*, 2023: 1-12
- [37] HODGKIN A L, HUXLEY A F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 1952, 117(4): 500
- [38] IZHIKEVICH E M. Simple model of spiking neurons. *IEEE Transactions on Neural Networks and Learning Systems*, 2003, 14(6): 1569-1572
- [39] ESHRAGHIAN J K, WARD M, NEFTCI E O, et al. Training spiking neural networks using lessons from deep learning. *Proceedings of the IEEE*, 2023, 111(9): 1016-1054
- [40] MASQUELIER T, THORPE S J. Learning to recognize objects using waves of spikes and spike timing-dependent plasticity// *Proceedings of the 2010 International Joint Conference on Neural Networks. Barcelona, Spain*, 2010: 1-8
- [41] SENGUPTA A, YE Y, WANG R, et al. Going deeper in

- spiking neural networks: VGG and residual architectures. *Frontiers in Neuroscience*, 2019, 13: 95
- [42] BELLEC G, SCHERR F, SUBRAMONEY A, et al. A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, 2020, 11(1): 3625
- [43] LIU Y H, SMITH S, MIHALAS S, et al. Cell-type – specific neuromodulation guides synaptic credit assignment in a spiking neural network. *Proceedings of the National Academy of Sciences*, 2021, 118(51): e2111821118
- [44] WU Y, DENG L, LI G, et al. Direct training for spiking neural networks: Faster, larger, better//Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii, USA, 2019: 1311-1318
- [45] FANG W, YU Z, CHEN Y, et al. Incorporating learnable membrane time constant to enhance learning of spiking neural networks//Proceedings of the IEEE/CVF International Conference on Computer Vision. Online, 2021: 2661-2671
- [46] STAN M I, RHODES O. Learning long sequences in spiking neural networks. *Scientific Reports*, 2024, 14(1): 21957
- [47] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141
- [48] KHALFAOUI-HASSANI I, PELLEGRINI T, MASQUELIER T. Dilated convolution with learnable spacings//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda, 2023: 1-29
- [49] WASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need//Proceedings of the 31st Annual Conference on Neural Information Processing Systems. California, USA, 2017: 1-11
- [50] WARDEN P. Speech commands: A dataset for limited-vocabulary speech recognition. *arXiv preprint arXiv:1804.03209*, 2018
- [51] CRAMER B, STRADMANN Y, SCHEMMEL J, et al. The heidelberg spiking data sets for the systematic evaluation of spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 33(7): 2744-2757
- [52] FANG W, CHEN Y, DING J, et al. Spikingjelly: An open-source machine learning infrastructure platform for spike-based intelligence. *Science Advances*, 2023, 9(40): eadi1480
- [53] KINGMA D P. Adam: A method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA, 2015: 7-9
- [54] SMITH L N, TOPIN N. Super-convergence: Very fast training of neural networks using large learning rates//Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications. Maryland, USA, 2019: 369-386
- [55] NOWOTNY T, TURNER J P, KNIGHT J C. Loss shaping enhances exact gradient learning with eventprop in spiking neural networks. *arXiv preprint arXiv:2212.01232*, 2022
- [56] DAMPFHOFFER M, MESQUIDA T, VALENTIAN A, et al. Investigating current-based and gating approaches for accurate and energy-efficient spiking recurrent neural networks//Proceedings of the International Conference on Artificial Neural Networks. Bristol, UK, 2022: 359-370
- [57] YIN B, CORRADI F, BOHTÉ S M. Accurate and efficient time-domain classification with adaptive spiking recurrent neural networks. *Nature Machine Intelligence*, 2021, 3(10): 905-913
- [58] BITTAR A, GARNER P N. A surrogate gradient spiking baseline for speech command recognition. *Frontiers in Neuroscience*, 2022, 16: 865897
- [59] SADOVSKY E, JAKUBEC M, JARINA R. Speech command recognition based on convolutional spiking neural networks//Proceedings of the 33rd International Conference Radioelektronika. Pardubice, Czech Republic, 2023: 1-5
- [60] HE X, LI Y, ZHAO D, et al. Msat: Biologically inspired multistage adaptive threshold for conversion of spiking neural networks. *Neural Computing and Applications*, 2024, 36(15): 8531-8547
- [61] DIEHL P U, NEIL D, BINAS J, et al. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing//Proceedings of the 2015 International Joint Conference on Neural Networks. Killarney, Ireland, 2015: 1-8
- [62] LEONTEV M, ANTONOV D, SUKHOV S. Robustness of spiking neural networks against adversarial attacks//Proceedings of the International Conference on Information Technology and Nanotechnology. Samara, Russia, 2021: 1-6



ZHANG Hang-Ming, Ph. D. candidate. His research interests are learning algorithm in spiking neural networks and neural coding.

BAI Qian-Yi, M. S. candidate. His research interests

focus on spike neuron learning algorithm and artificial intelligence.

DENG Zhi-Chao, M. S. candidate. His research interests focus on brain-inspired computing and neural networks.

Alexander Sboev, Ph. D., professor. His research interests focus on neuromorphic computing and memristive plasticity.

YU Qiang, Ph. D., professor. His research interests include brain-inspired intelligence and computing, neural networks, artificial intelligence, and theoretical neuroscience.

Background

Spiking Neural Networks (SNNs) have garnered significant attention in recent years due to their low energy consumption, event-driven characteristics and robustness to noise. Leveraging their efficient processing of temporal information, SNNs have shown promising applications in temporal tasks such as speech recognition. It has been shown that the synaptic plasticity delays mechanism plays a significant role in enhancing the temporal modeling ability and performance of SNNs, but the existing synaptic plasticity delays learning methods still have limitations in dealing with the noise interference, data sparsity, and inhomogeneity in the real-world environments, which makes SNNs less effective for complex tasks. Therefore, optimizing the existing synaptic plasticity delays learning methods to enhance its ability to adapt to complex environments and capture temporal features is one of the key directions to improve the performance of SNNs.

To address this issue, we introduce the temporal attention mechanism into the synaptic plasticity delays learning method. This approach enhances the attention of SNNs to key timesteps

by introducing the attention mechanism to more accurately capture essential information in the input signal. First, we utilize the dilated convolution with learnable spacings to learn and optimize delay parameters and their weights. Then we introduce a squeezing operation to reduce the complexity of data processing. Next, we further introduce a one-dimensional convolution-based temporal attention mechanism that enables SNNs to pay more attention to the timesteps in the data that contain critical information. Experimental results show that this method exhibits significant advantages in speech recognition tasks. In addition, we systematically explore different ways of combining the two mechanisms, and the results show that placing the temporal attention mechanism between the synaptic plasticity delays and the spiking layer can capture the key features more effectively, which provides an experimental basis for the design and optimization of network architectures.

This research was supported by the National Natural Science Foundation of China (Grant Nos. 92370103, 62176179) and the Xiaomi Young Scholar Program.