

专家观点



[专家简介] 王卫东,解放军总医院医学创新研究部某中心主任,博士,研究员,博士生导师,北京大学、北京理工大学、北京航空航天大学、中国科学院大学兼职教授。长期从事人工智能、信号与信息处理、生物医学工程等研究和医疗器械研发。负责多项国家科技创新重大项目、国家重点研发计划,担任国家战略性新兴产业发展专家咨询委员会委员、中国生物技术发展中心专家委员会委员、中国生物工程学会生命科学仪器专业委员会副主任委员。曾任中国医学装备协会医学装备技术保障专业委员会副主任委员、中国计量测试学会医学计量分会副主任委员等职务。获军队科技进步一等奖1项、黄家驷生物医学工程奖一等奖1项,享受国务院政府特殊津贴,授权国内专利、软件著作权30多项。

感觉编码时空活动模式的类脑表达与计算——算法篇

王卫东^{1,2}

¹解放军总医院医学创新研究部,工业和信息化部生物医学工程与转化医学重点实验室,北京 100853;

²北京航空航天大学生物与医学工程学院,北京 100191

摘要:在深入理解视网膜、外侧膝状体及大脑视皮质之间的视觉编码神经生理机制的基础上,我们构建了一种新型的类脑表达和计算的神经网络单元架构,为大规模集成神经网络的类脑大模型实现奠定了基础。这一类脑智能单元架构,为我们理解大脑感觉编码的表达和计算提供了信息论基础。本文进一步探讨了类脑智能单元及其大模型的训练方法、策略以及具体的算法实例,提出了一种综合性策略。该策略结合了感觉数据流表达与计算的冗余减少原则、自组织映射以及折返振荡同步机制,能够实现大脑全局编码,旨在提升类脑大模型的生物合理性和可解释性,以及高效快捷地模仿行为控制和意识等复杂的大脑功能。

关键词:冗余减少;压缩编码;稀疏编码;自组织映射;脑连接图;脑科学;类脑表达

中图分类号:R338.2;TP18

文献标志码:A

文章编号:2095-5227(2025)01-0008-08

DOI: 10.12435/j.issn.2095-5227.2024.111

引用本文:王卫东.感觉编码时空活动模式的类脑表达与计算——算法篇 [J].解放军医学院学报, 2025, 46 (1): 8-15.

Brain-inspired representation and computation for similarity structure from spatiotemporal patterns in sensory coding: Algorithms

WANG Weidong^{1,2}

¹Medical Innovation Research Department of PLA General Hospital, Key Laboratory of Biomedical Engineering and Translational Medicine, Ministry of Industry and Information Technology, Beijing 100853, China; ²School of Biological and Medical Engineering, Beihang University, Beijing 100191, China

Corresponding author: WANG Weidong. Email: wangwd301@126.com

Abstract: Based on a deep understanding of the visual encoding neurophysiological mechanisms between the retina, lateral geniculate body, and visual cortex of the brain, we have constructed a novel neural network-based brain-inspired intelligent unit architecture, laying a solid foundation for the implementation of large-scale integrated neural network-based brain-inspired models. This information theory foundation for our understanding of the expression and computation of brain sensory encoding is formed by the architecture of brain-inspired intelligent units. This article delves into the training methods, strategies, and specific algorithm

收稿日期: 2024-04-17

基金项目: 新一代人工智能国家科技重大专项(2020AAA0105801)

通信作者: 王卫东, 博士, 研究员, 博士生导师. Email: wangwd301@126.com

examples of brain-inspired models, and proposes a comprehensive strategy that combines the redundancy reduction principle of sensory data flow expression and computation, self-organizing feature mapping, and recurrent oscillation synchronization mechanism, aiming to improve the biological rationality and interpretability of brain-inspired models, as well as efficiently and quickly mimic complex brain functions.

Keywords: redundancy reduction; compression coding; sparse coding; self-organization mapping; brain connectivity map; brain science; brain-inspired representation

Cited as: Wang WD. Brain-inspired representation and computation for similarity structure from spatiotemporal patterns in sensory coding: Algorithms[J]. Acad J Chin PLA Med Sch, 2025, 46(1): 8-15.

受大脑神经生理学机制的启发,我们构建了基于联结主义机制的表达与计算的类脑智能单元架构,这一架构展现出卓越的生物学合理性和可解释性。在此基础上,建立该架构的数学模型和信息论理论基础显得尤为关键。鉴于物种所处的多样化生态环境中自然刺激信号的高度信息冗余性,单纯依赖感官细胞进行信号的时空采样表达,往往难以达到智能处理、快速响应及高效利用的目标。因此,感官通路在早期阶段主要致力于重新编码接收到的信号,将其转化为更高效、更利于泛化的表达形式,即感受野和投射野的表达方式。这种编码策略不仅促进了表达学习的进程,也推动了认知计算和行为控制的实现。在信息表达学习的过程中,高效的感官编码始终保持着信息冗余减少与信息熵增大之间的平衡。目前,构建神经编码表达和计算的数学理论已成为研究人员关注的焦点。

Barlow^[1-2]最先提出了感觉编码的信息冗余减少假说,2001年Barlow又重新审视并修正该假说,将过严的冗余减少策略变成了一种宽松的冗余减少^[3],做到了冗余与熵的信息度量之间的平衡,从而通过适当地保留一些冗余,以削弱噪声带来的负面影响,被人们称为感觉编码的第一性原理^[4]。Barlow对冗余与感觉和知觉的神经生理机制之间的关系更感兴趣,Attneave^[5]更关心知觉方面,而Watanabe^[6]则注重高层次推理,但他们又一致认为来自自然环境的物理刺激具有冗余性。在Barlow看来,其先前的冗余减少假说过分强调压缩编码以及仅关注对神经元数量的作用是错误的,而强调冗余的重要性以及在先验条件下的冗余减少才是正确的学说。针对基于感觉编码第一性原理形成的稀疏编码^[7-8]、盲源分离^[9-10]、独立成分分析^[11-13]、因式分解编码^[14-15]、熵最小化编码^[16]、信息最大化编码^[17-18]、典型相关分析^[19]等方法,Nadal和Parga^[20]阐明了它们之间的联系和区别,以及在非线性输出的低突触噪声极限中,当对突触强度和非线性传递函数进行优化时,信息最大化

意味着冗余减少。人们在研究丘脑与皮质之间如何改变感觉信号的空间和时间特征以及如何调节信号传输的过程中,创立了因式分解编码和信息熵最小化的计算框架,即将感官信息的概率分布函数分解成最小信息熵成分的概率分布函数乘积^[4,14]。这种因式分解编码的无监督学习方法能够找到独立视觉特征,与猫和猴的初级视觉皮质中发现的具有线段和边缘选择性刺激-反应的神经元形成的自然场景独立、稀疏、分布表达非常类似。而盲源分解和独立成分分析的计算框架,则明确了感官信息被相互独立的感受野分离成独立成分编码^[21-23]。Mallat^[24]提出的字典学习与稀疏编码的计算框架,强调感官信息一次只能使用少量潜在活跃神经元进行编码^[25-26],即寻找在冗余字典或转换上的稀疏表达。Linsker^[27]创立了互信息最大化的计算框架,其中感官信息被编码前后的互信息最大可能地保留有用的信息。这种基于信息最大化的无监督学习方法,也是一种非线性信息最大化网络,应用于自然场景的集合时,会产生一组局部、取向的视觉滤波器,类似于稀疏最大化网络产生的Gabor滤波器。此外,这些滤波器的输出尽可能独立,因为该网络对稀疏分量分布执行独立分量分析,相比基函数与主成分分析和零相位白化滤波器实现的去相关具有在自然场景中更稀疏分布的输出。理论上,因式分解编码^[28]与冗余减少原理、盲源分离与独立成分分析、冗余字典生成与稀疏编码具有共同的理论基础和相近的计算结果,均实现了感官世界信息统计学信号数据流的相关性最小化,形成独立成分的超完备稀疏表达,也类似于视觉皮质中简单细胞的感受野,这表明这些神经元形成了一个信息学意义上的自然图像坐标系。

近年来利用深度神经网络开展对比监督学习或自监督学习的算法研究逐渐受到重视,这些算法本质上是对深度学习网络中的数据流进行冗余度消除、信息熵降低、独立成分提取、稀疏表达输出处理。最早提出的自监督学习孪生网络^[29],

采用标准交叉熵损失或均方差损失作为优化准则来评估数据及其增强数据之间的相似性。通过时间与空间变化的数据增强在孪生网络变体深度学习中有广泛应用，如 SimCLR、SimSIAM、BYOL、VICReg 和 SwAV 等使用数据增强技术在网络对中引入不对称，这些变体均采用了标准交叉熵损失或均方差损失作为对比相似性或不变性评价准则^[30-34]。通过输入数据流的滤波、内插、预测、动态差分、积分、平滑、分割、分段、分时和掩膜等运算获得对比数据流或自监督数据流，这些图像和信号的空间失真和时间变化的数据增强预处理方法，摆脱了监督学习需要大量人工标识数据的困境。通过数据流在时间和空间域分割单元之间相关编码实现数据流的表达学习，寻找数据流中时空模式的相似性结构，即从样本或数据流的变化中找到不变的变换网络，感觉编码上等同于统计学意义上相互独立的非高斯成分。最近，Jure 等^[35]提出的 Barlow 孪生网络采用了不同的对比相似性评价准则，引入冗余减少概念，即 Barlow 冗余减少原理^[3]。但经过感觉通路处理、加工和表达的信息进入大脑皮质进一步综合利用，远非那么简单。我们基于视网膜到丘脑外侧膝状体再到大脑皮质的网络结构和信息转换机制，构建了一种面向表达和计算的感觉神经编码类脑智能单元模型。在此模型中，压缩-稀疏编码网络担任主干网络的核心角色，而单元间的接口网络则巧妙地作为时间和空间上下文承接层投射网络，实现了信息的有效传递与整合。此外，模型还引入自组织映射网络作为单元计算内核层，该网络之间通过折返、耦合、振荡网络实现了信息的交互与协同，不仅丰富了模型的计算层次，也增强了其对复杂信息的处理能力。

本文提出了基于冗余减少原则的主干网络训练算法，该算法通过计算参考数据集及其时空变化、样本集之间的正交投影误差或稀疏表达误差，实现误差反向传播，从而优化网络性能。而自组织映射网络和折返、耦合、振荡网络的训练过程可独立于主干网络进行，大大提高了训练的灵活性。在模型训练过程中，我们综合运用深度学习、自组织映射、稀疏编码、压缩编码等算法，进一步提升了类脑智能单元模型的准确性和泛化能力。

1 类脑智能单元架构的学习框架

在感觉信息处理中，类脑表达与计算遵循一

个核心原理：在丘脑中，信息通过系统分离和近邻排列的神经元协同激活来精准表达瓶颈信息，即所谓的压缩编码。而在大脑皮质中，这一过程则由相对较少的活跃神经元在大量神经元中脱颖而出表达信息，即稀疏编码。压缩编码的神经元和稀疏编码的神经元能够自组织映射并进行拓扑计算，而不同自组织映射表达层之间的折返连接则形成大脑皮质的振荡及其同步全局编码表达的脑功能地形图。在进行单元学习时，我们采取了模块化的训练策略，将其分解为3个相对独立的部分进行训练。这样做的目的是确保瓶颈信息在后续深度皮质处理中具有更强的泛化能力。从信息处理的前馈视角来看，外周神经系统首先对感觉信号进行系统性分离与压缩编码，随后这些信号通过神经纤维高效地传递到丘脑。在丘脑中，压缩后的信号经过近邻排列与合并融合，进一步精炼，随后通过基底神经节稀疏地传递到大脑皮质进行高级处理。数据流中的时间-空间活动模式相似性结构(即时空结构)极为重要，是人类行为的基石，也是类脑计算元架构中时空结构表达问题的关键所在。我们并未直接在时空结构中显式表达，而是巧妙地通过处理过程的影响来隐含地表达时空结构，这种策略不仅提升了信息处理的效率，还确保了处理的精准性。

考虑时间-空间上下文的类脑智能单元模型的冗余减少数学模型如图1所示。不失一般性， N 个单元的空间压缩表达层的压缩函数分别 $f_i(I_i) \in R^n, i = 1, 2, \dots, N$ ，其中 $I_i \in R^k$ 为第 i 个类脑模型单元的输入， $f_i(*)$ 是一个可以训练的深度神经网络，如卷积神经网络、残差神经网络、自注意力网络。 $O_i \in R^n$ 为上时刻第 i 个计算单元的输出，则此时刻的第 i 个计算单元的压缩表达层输出 $y_i \in R^n$ 为：

$$y_i = f_i(I_i) + w_{ii}O_i + \sum_{j=N_1}^{N_2} w_{ij}O_j, i = 1, 2, \dots, n \quad (1)$$

这里 $w_{ij} \in R^{n \times n}$ 是第 i 个单元 N_i 个近邻单元(从 N_1 到 N_2)空间上下文的投影矩阵， $w_{ii} \in R^{n \times n}$ 是第 i 个单元时间上下文的投影矩阵，且压缩表达层输出 y_i 存入下时刻第 i 个类脑计算单元的输出 $O_i \in R^n$ ，即将压缩表达层输出寄存起来 $y_i \xrightarrow{\text{存储}} O_i$ 。(1)式就是所谓的 Elman 递归神经网络。

构建类脑智能单元模型压缩表达层输出信号 y_i 的稀疏和冗余模型，将其表达为来自字典

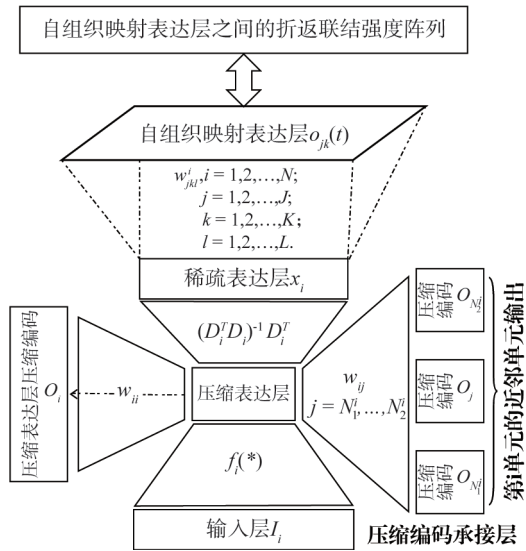


图1 时间-空间上下文的类脑智能单元模型的数学模型

Fig. 1 A mathematical model of a brain-inspired intelligence unit model in a spatiotemporal context

$D \in R^{n \times m}$ 的几个信号的线性组合。这个字典包含 m 个原子 $d_j \in R^n$ 的集合，是稀疏表达的构建单元。信号 y_i 的表达是任意矢量 $x_i \in R^m$ ，满足 $y_i = D_i x_i$ 。

在 $m > n$ 的情况下，稀疏表达层输出信号 x_i 是冗余的，即系统 $y_i = D_i x_i$ 有无穷多个可能解。在这个解集中，首选最稀疏的解，即具有最小的 $\|x_i\|_0$ 范数解。因此，计算信号 y_i 的稀疏表达层信号 x_i 的问题可以描述为：

$$\hat{x}_i = \arg \min_{x_i} \|x_i\|_0, \text{ s.t. } y_i = D_i x_i \quad (2)$$

约束条件放宽为 $\|y_i - D_i x_i\|_2 \leq \varepsilon$ ，以适用于附加噪声和模型偏差，则(2)式精确等价于：

$$\hat{x}_i = \arg \min_{x_i} \|y_i - D_i x_i\|_2, \text{ s.t. } \|x_i\|_0 \leq L \quad (3)$$

其中 $L \in [1, 2, \dots, m-1]$ 称为稀疏度。该问题的求解被证明是一个 NP 难题。目前常用匹配追踪方法和基追踪算法进行近似求解。理论研究表明，如果存在一个足够稀疏的解，则可以通过这些近似方法找到它。要成功地将上述模型应用到信号处理中，关键是字典 D_i 的选择。尽管基于小波、小波包、可引导小波、曲线小波、轮廓小波和带小波预定义字典的方法广泛使用，但其并不适用于类脑表达与计算模型。

适用于类脑表达和计算模型单元的训练技术是设计字典，构建一个大小为 $n \times m$ 的字典 D_i ，实现训练信号的最稀疏表达，即通过学习过程有效地表达给定的信号集。将信号的稀疏和冗余表达建模与学习字典相结合，在纹理分类、静态图像

和视频去噪、彩色图像修补和去噪等图像处理方面具有很大的优势。

1.1 稀疏表达误差反向传播算法

本文通过数据集及其时空失真压缩编码的冗余字典 D_i 学习，获得它们的稀疏表达系数 \hat{x}_i 和 \hat{x}_i' ，然后计算出它们的压缩表达层信号 y_i 和 y_i' 的估值 \hat{y}_i 和 \hat{y}_i' ，以及误差 $(\hat{y}_i - \hat{y}_i')$ ，最后通过误差反向传播算法训练并获得压缩表达网络 f_i^* 和时间-空间上下文投影连接矩阵 w_{ij} 。基于稀疏表达技术和感觉编码冗余减少原则的类脑表达学习和计算算法如下。

预处理：选择感觉刺激输入信号 I_i (或一批)，初始化 O_i 、 f_i^* 和 w_{ij} ，以及 \hat{x}_i 和 D_i ，根据感觉刺激输入信号 I_i 计算压缩表达信号 y_i ，计算 y_i 的稀疏编码 \hat{x}_i ；第 1 步，感觉刺激输入信号 I_i (或一批) 进行失真变换获得 I_i' (或多个失真变换)；第 2 步，根据感觉刺激输入信号 I_i 及其失真 I_i' 计算压缩表达信号 y_i 及其失真压缩表达信号 y_i' ；第 3 步，计算 y_i 和 y_i' 的稀疏编码 \hat{x}_i 和 \hat{x}_i' ；第 4 步，进行冗余字典 D_i 学习迭代计算；第 5 步，判别稀疏编码是否收敛，“是”则计算压缩表达信号 y_i 和 y_i' 的估值 $\hat{y}_i = D_i \hat{x}_i$ 和 $\hat{y}_i' = D_i \hat{x}_i'$ ，“否”则跳转第 3 步；第 6 步，计算压缩表达层信号 y_i 的误差 $(\hat{y}_i - \hat{y}_i')$ ；第 7 步，利用误差反向传播算法训练 O_i 、 f_i^* 和 w_{ij} ；第 8 步，判别压缩编码是否收敛，“是”则选择新的感觉刺激输入信号 I_i (或新的一批) 跳转第 2 步，“否”则根据感觉刺激输入信号 I_i 及其失真 I_i' 计算压缩表达信号 y_i 及其失真信号 y_i' 后跳转第 6 步。

我们提出的基于新型冗余减少第一性原理的类脑表达和计算的自学习训练方法，是一种无监督学习技术，即通过独立成分分析或因式分解或互信息最大化等方法，实现压缩表达层信号 y_i 及其失真 y_i' 的估值 \hat{y}_i 及其失真 \hat{y}_i' 计算，然后通过数据集及其失真样本集的压缩表达层信号估值误差 $(\hat{y}_i - \hat{y}_i')$ 反向传播算法训练并获得 f_i^* 和 w_{ij} 。值得强调的是，采用独立成分分析和冗余空间投影误差同样可以实现上述类脑表达和学习训练方法。

1.2 皮质自组织特征映射算法

在类脑模型元架构中，用于神经感觉信号表达和计算的各类脑单元能够独立训练，也可以基于无标签数据集利用无监督学习算法进行预训练。神经表达和计算类脑单元的神经元可以选用数字神经元、模拟神经元和脉冲神经元。压缩表达层输出信号采用与预训练相同的数据集进行自组织映射，即训练类脑模型元架构中自组织映射网络阵列，自组织网络阵列的不同单元可以采用经典的 Kohonen 自组织映射算法独立进行训练。Kohonen 算法本质上是一种具有侧向抑制特性、保持数据拓扑结构不变、矢量量化与矢量符号化的神经网络，通过赢者通吃型侧向抑制连接权重调节法则。不失一般性，第 i 个类脑单元自组织映射阵列单元的 Kohonen 训练算法可描述如下：

第 1 步，初始化： $w_{jkl}^i, i = 1, 2, \dots, N; j = 1, 2, \dots, J; k = 1, 2, \dots, K; l = 1, 2, \dots, L$ ；第 2 步，选择感觉刺激输入信号 I_i （或一批），根据感觉刺激输入信号 I_i 计算压缩表达信号 $y_i = [y_{i1} y_{i2} \dots y_{iL}]^T$ 并转化成 Poisson 分布编码的脉冲串；第 3 步， $Y(t) = [i_1(t) i_2(t) \dots i_L(t)]^T, t = 0, 1, \dots, N_t - 1$ ，其中 $i_1(t), i_2(t), \dots, i_L(t)$ 为输入层 t 时刻输入的脉冲零或幺；第 4 步，计算自组织映射输出层第 j, k 个神经元的输出脉冲 $o_{jk}(t), t = 0, 1, \dots, N_t - 1$ ；第 5 步，

计算放电速率 $r_{jk} = \frac{\sum_{t=0}^{N_t-1} o_{jk}(t)}{N_t}$ ， r_{jk} 最大者为胜者；

第 6 步，按照 Kohonen 自组织映射算法更新 w_{jkl}^i ，转到第 2 步。该算法的压缩表达层输出信号为模拟信号，也可以选择产生脉冲信号的脉冲神经网络，如果选择脉冲神经网络就无须将模拟信号转化成 Poisson 分布编码的脉冲信号串。同样，上述算法也适用于稀疏编码表达层输出的自组织映射。

在自组织映射阵列训练的基础上，进行自组织映射阵列之间折返联结动态连接强度的训练。值得强调的是，采用独立成分分析同样可以实现上述基于稀疏编码和新型冗余减少原则的类脑表达和学习训练方法。

1.3 皮质折返联结振荡同步算法

在自组织映射表达层阵列训练的基础上，进一步开展其阵列之间振荡与同步振荡折返连接网络训练，这种符号化的类脑振荡与同步振荡形成连接网络图，对不同任务的脑连接图实现全局通信、关联和任务捆绑等。图 2 给出了实现脑区网

络与皮质网络之间振荡和同步振荡的方式，这种振荡和同步振荡通过自组织网络之间的动态连接强度训练实现。图 2 中的不同功能皮质之间采用折返联结的动态连接权重，实现皮质之间的脉冲耦合同步振荡，从而实现皮质之间的通信、关联、相互激活和镜像，形成皮质功能动态连接网络图。不同功能皮质之间采用折返联结的动态连接权重，实现皮质之间的脉冲耦合同步振荡，从而实现皮质之间的通信、关联和镜像，构建图连接方式的脉冲神经网络，实现贝叶斯网络的符号主义联结机制，为概率图模型的推理决策提供联结主义可能性。设一个自组织映射表达层阵列阵元内的神经元 i 与另一个自组织映射表达层阵列阵元内的神经元 j 之间的动态联结权重为 $w_{ij}^{act}(t+1)$ ，则该动态联结权重的训练可采用如下算法：

$$w_{ij}^{act}(t+1) = w_{ij}^{act}(t) + R \cdot \delta \cdot \text{sigmoidal}(w_{ij}^{act}(t)) \cdot (\bar{s}_i(t) - \theta_i) \cdot (\bar{s}_j(t) - \theta_j) \quad (4)$$

这里 $w_{ij} = w_{ij}^{rest} + w_{ij}^{act}$ ，平均场 $\bar{s}_{ij}(t)$ 是脑区脉冲神经元膜电位的总和平均值，即 $\bar{s}_{ij}(t) = \frac{t-1}{t} \bar{s}_{ij}(t-1) + \frac{1}{t} s_{ij}(t)$ ， $\bar{s}_{ij}(0) = s_{ij}(0)$ 。其中 $s_{ij}(t)$ 是自组织映射表达层网络两个不同单元中第 i, j 个神经元的活性，计算如下：

$$s_i(t) = s_i(t-1) + \theta \left(\sum_k w_{ik} s_k(t) + \sum_l w_{il}^{in} x_l \right) + N \quad (5)$$

$$s_j(t) = s_j(t-1) + \theta \left(\sum_m w_{jm} s_m(t) + \sum_n w_{jn}^{in} x_n \right) + N \quad (6)$$

这里 $w_{ik,jk} = w_{ik,jm}^{rest} + w_{ik,jm}^{act}$ 。 w_{ik} 和 w_{jm} 表示第 i, j 神经元的近邻神经元侧向连接权重， w_{il}^{in} 和 w_{jn}^{in} 表示第 i, j 神经元的外部压缩表达层输入连接权重， x_l 和 x_n 表示第 i, j 神经元的外部输入信号。如果 $\bar{s}_i(t) \geq \theta_i$ 或 $\bar{s}_j(t) \geq \theta_j$ ，那么 $R = 1$ ，否则 $R = 0$ 。

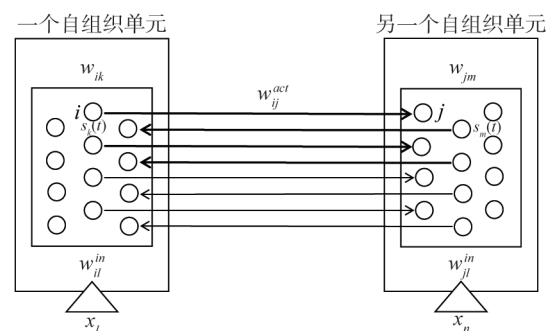


图 2 自组织之间的折返连接耦合振动的数学模型
Fig. 2 A mathematical model of the reentrant connection coupling vibrations between self-organizing units

基于冗余减少准则的预训练技术、侧向抑制的自组织映射机制以及振荡与同步振荡图网络的自学习能力,我们构建了一个统一的表达和计算框架,该框架能够支持多种智能应用任务,包括识别、跟踪和控制等。在这个框架中,压缩表达层负责提取感觉信息流的特征表达,而自组织映射表达层则输出振荡与同步振荡图网络。这些图网络不仅实现了特征的分解和捆绑,而且能够利用压缩表达层的输出作为深度神经网络的输入。利用标签数据集和监督学习算法,我们能够对这些深度网络进行精细训练。最终,通过联合自组织映射阵列的连接网络图,我们能够实现广泛的人工智能应用,从而满足多样化的任务需求。

2 感官信号的类脑智能单元模型训练的讨论

在感知通路中,压缩表达层的感受野与投射野、稀疏表达层的自组织映射及其之间的折返联结,作为对自然环境中纷繁复杂的刺激进行高效处理的特殊信号处理机制,发挥着至关重要的作用。然而,感受野、投射野、振荡及其同步的大脑皮质全局编码机制,至今仍是神经科学领域亟待揭示的问题,我们所构建的训练算法为上述机制的揭示提供了可能的理论支持。在自然环境中,感觉刺激信号的特征是冗余的,因此单纯依赖感觉细胞阵列进行信号采样表达的方式效率很低。相较之下,采用感受野和投射野的方式进行信号采样表达则更为高效,并有助于推动学习过程的进行。从进化论和信息论的双重视角来看,丘脑皮质间的感受野与投射野在信号重新编码后,不仅展现出进化论上的生存遗传优势,还凸显了信息论上的高效运用优势。这些优势使得大脑皮质能够对信号进行多层次、更深入的智能化计算和重复应用开发,从而在竞争激烈、机会驱动的环境中做出明智决策。在这一过程中,大脑必须能够精准地理解和利用非随机概率,以及感官物体和事件之间的相互依赖性。

本质上,冗余减少是一种无监督学习策略,它通过去除重复的数据、消除数据的相关性、提高数据的信息熵等对感官数据的冗余信息进行压缩,在泛化地获得数据特征的同时,可以有效减少数据量和计算量,实现数据的随机性和非随机性信息分离。Barlow和Tolhurst^[36]提出视觉皮质特征探测器可能是冗余减少过程的最终结果,其认

为每个特征探测器的激活尽可能与其他探测器在统计上独立。Atick等^[4]进而提出了感觉信号处理的信息论生态理论,并由Hubel和Marr等^[37-38]证实。稀疏编码、独立成分分析、信息最大化网络、因式分解编码、孪生网络等冗余减少策略的理论方法和无监督学习算法,很好地诠释了视觉皮质中简单细胞的感受野形成是自然场景的成分独立化、信息最大化的稀疏表达的信息理论,这些都融合了Barlow的冗余减少信息学原则。Zylberberg等^[39]在数学上进一步证明,使用突触局部可塑性规则来学习稀疏性和去相关是优化合作线性生成图像模型的关键,可以弥合视觉皮质的理论和实验研究之间的差距。最近提出的Barlow-Twins等孪生网络虽然也是基于Barlow冗余减少原则的自监督学习算法,但生物学解释性较差。本文提出的方法采用自组织映射、稀疏编码和字典学习、分组与组合动态学习及其稀疏解码、误差反向传播训练压缩编码网络的联合算法,实现了新型生物学可解释的冗余减少信息学原则。现有的基于冗余减少的类脑建模理论方法很少考虑外周神经系统和丘脑对感官数据的压缩编码和拼接融合的行为功能,均采用了感官数据直接稀疏化、成分去相关或独立化、概率分布因式分解和信息最大化编码等来实现。它们之间具有等价关系,使得每个特征探测器被尽可能独立激活、统计上独立成分分析意味着信号分离具有非线性信息最大化能力^[40-41]。信息最大化意味着概率密度函数的因式分解编码,它可能涉及所有阶的依赖性,但大多数基于Barlow冗余减少原理的理论方法研究仅使用了一组分布式系数局部特征检测器输出所需的二阶统计信息去相关,就能解释早期视觉的心理学和神经生理学特性,且实现冗余和维度的同时减少,以非高斯方式构建整体局部特征编码的方法远低于奈奎斯特采样率^[42-44]。

就算法而言,许多实用算法具有相似性,且没有完全包含冗余减少原理的深刻内涵。就结构而言,这些算法没有考虑丘脑在冗余减少中的作用,即对感觉信号进行感受压缩和拼接融合,从而形成更通用、一般化和泛化的潜在表达^[45]。就集成而言,潜在表达能够通过折返连接方式将大脑皮质中的计算单元整合起来,成为大规模集成神经网络模型的基本单元及核心信息表达方式^[46]。为此,本文重点研究了对感觉信号进行降低冗余度、去除相关性、减小信息熵、增加噪声熵、最

大化互信息的压缩编码表达,并将不同单元的压缩编码整合后独立稀疏地投射到大脑皮质,形成统计上尽可能独立的稀疏表达的感受野或功能柱,以此实现概率推理的神经网络结构、算法及其集成方式^[47]。

现有基于冗余减少策略的类脑数学建模训练理论与方法,未考虑外周神经系统编码与丘脑的隐含压缩和拼接融合在表达学习与智能计算中的核心作用,也未系统地考虑大脑皮质的自组织和稀疏表达在计算智能中的作用。因此,如何将它们的运行机制有效地综合应用于自学习和预训练的类脑建模中,是非常值得研究的基础科学问题。至今,涉及其核心研究的相关报道甚少。本文提出的基于冗余减少原理的感觉信息表达和计算的理论框架及方法算法,引入了具有数据压缩和池化计算特性的压缩表达层,它是采集感觉数据的感觉器官的感受野表达;引入了具有稀疏/独立/因式分解/最大化计算特性的稀疏表达层,它是压缩层的投射野表达。由于我们的框架和方法采用了上下文拼接融合,使得自学习、预训练、数据产生和大规模神经网络模型集成成为可能。这充分彰显了在大脑表达学习中丘脑数据隐含压缩和上下文拼接的功能作用,其所习得的表达记忆和长期记忆被存储在丘脑中,而自组织学习和稀疏表达由皮质完成。从感觉器官到丘脑的压缩表达,再到感觉皮质的自组织学习和稀疏表达,在通过本文提出的预训练自监督学习算法完成的同时获得了隐含压缩表达,其信息可用于其他皮质的深层智能计算,即通过振荡和同步振荡相关的短时记忆、工作记忆、多时钟计算形成认知智能和行为控制。

3 结论

本文提出的基于类脑智能单元的大规模集成神经网络模型的学习算法对理解大脑的生物学结构与功能、大脑的信息学计算与表达,以及发展脑科学和类脑技术具有潜在的应用价值和意义。

大规模集成神经网络是电子与信息技术领域的里程碑式事件,依据大脑感觉编码的生物学机制,构建结构化、模块化、集成化的大规模类脑智能模型,是创建强人工智能的必由之路。本文提出了基于Barlow冗余减少原理的稀疏编码或正交投影误差反向传播算法,引入自组织映射和折返耦合振荡网络的训练算法,解决了类脑智能单

元及其组建的大模型训练方法,有望推动神经现象学中的振荡及其同步应用于表达和计算。

利益冲突 作者声明无利益冲突。

参考文献

- 1 Barlow HB. Sensory mechanisms, the reduction of redundancy, and intelligence [C] //Proceedings of the symposium on the mechanization of thought processes. London: National Physical Laboratory, 1959.
- 2 Barlow HB. Possible principles underlying the transformations of sensory messages [M/OL]. <https://doi.org/10.7551/mitpress/9780262518420.003.0013>.
- 3 Barlow H. Redundancy reduction revisited [J]. *Network*, 2001, 12 (3): 241-253.
- 4 Atick JJ, Li ZP, Redlich AN. Understanding retinal color coding from first principles [J]. *Neural Comput*, 1992, 4 (4): 559-572.
- 5 Attneave F. Some informational aspects of visual perception [J]. *Psychol Rev*, 1954, 61 (3): 183-193.
- 6 Watanabe S. Information theoretical analysis of multivariate correlation [J]. *IBM J Res Dev*, 1960, 4 (1): 66-82.
- 7 Zhao HL, Li ZN, Su WS, et al. Dynamic sparse coding-based value estimation network for deep reinforcement learning [J]. *Neural Netw*, 2023, 168: 180-193.
- 8 Bowren J, Sanchez-Giraldo L, Schwartz O. Inference via sparse coding in a hierarchical vision model [J]. *J Vis*, 2022, 22 (2): 19.
- 9 Tang MY, Wu YF. A blind source separation method based on bounded component analysis optimized by the improved beetle antennae search [J]. *Sensors*, 2023, 23 (19): 8325.
- 10 Ansari S, Alatrany AS, Alnajjar KA, et al. A survey of artificial intelligence approaches in blind source separation [J]. *Neurocomputing*, 2023, 561: 126895.
- 11 Hyvärinen A, Khemakhem I, Morioka H. Nonlinear independent component analysis for principled disentanglement in unsupervised deep learning [J]. *Patterns (N Y)*, 2023, 4 (10): 100844.
- 12 Mestre-Bach G, Granero R, Fernández-Aranda F, et al. Independent component analysis for Internet gaming disorder [J]. *Dialogues Clin Neurosci*, 2023, 25 (1): 14-23.
- 13 Sreevalsan-Nair J. Independent component analysis [M/OL]. https://doi.org/10.1007/978-3-030-26050-7_158-1.
- 14 Hamamoto R, Takasawa K, Machino H, et al. Application of non-negative matrix factorization in oncology: one approach for establishing precision medicine [J]. *Brief Bioinform*, 2022, 23 (4): bbac246.
- 15 Aonishi T, Maruyama R, Ito T, et al. Imaging data analysis using non-negative matrix factorization [J]. *Neurosci Res*, 2022, 179: 51-56.
- 16 Wu XF, Cheng L, Zhang SF. Open set domain adaptation with entropy minimization [C] //Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Cham: Springer, 2020: 29-41.
- 17 Poduval P, Oberoi G, Verma S, et al. BipNRL: mutual information maximization on bipartite graphs for node representation learning [C] //Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2023: 728-743.
- 18 Wan WQ, Chen JL, Xie JS. MIM-Graph: a multi-sensor network approach for fault diagnosis of HSR Bogie bearings at the IoT edge via mutual information maximization [J]. *ISA Trans*,

- 2023, 139: 574-585.
- 19 Wróbel S, Turek C, Stepień E, et al. Data integration through canonical correlation analysis and its application to OMICS research [J]. *J Biomed Inform*, 2024, 151: 104575.
- 20 Nadal JP, Parga N. Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer [J]. *Netw Comput Neural Syst*, 1994, 5 (4): 565-581.
- 21 Młynarski WF, Hermundstad AM. Efficient and adaptive sensory codes [J]. *Nat Neurosci*, 2021, 24 (7): 998-1009.
- 22 Kafashan M, Jaffe AW, Chettih SN, et al. Scaling of sensory information in large neural populations shows signatures of information-limiting correlations [J]. *Nat Commun*, 2021, 12 (1): 473.
- 23 Akhavan S, Soltanian-Zadeh H. Blind separation of sparse sources from nonlinear mixtures [J]. *Digit Signal Process*, 2021, 118: 103220.
- 24 Mallat S. Understanding deep convolutional networks [J]. *Philos Trans A Math Phys Eng Sci*, 2016, 374 (2065): 20150203.
- 25 Romyantsev OI, Lecoq JA, Hernandez O, et al. Fundamental bounds on the fidelity of sensory cortical coding [J]. *Nature*, 2020, 580 (7801): 100-105.
- 26 Steinfeld R, Tacão-Monteiro A, Renart A. Differential representation of sensory information and behavioral choice across layers of the mouse auditory cortex [J]. *Curr Biol*, 2024, 34 (10): 2200-2211.
- 27 Linsker R. Self-organization in a perceptual network [J]. *Computer*, 1988, 21 (3): 105-117.
- 28 Wu YN, Sai GL, Duan SY. Work-in-progress: accelerated matrix factorization by approximate computing for recommendation system [C] // 2022 International Conference on Embedded Software (EMSOFT). IEEE, 2022: 1-2.
- 29 Bromley J, Guyon I, LeCun Y, et al. Signature verification using a "Siamese" time delay neural network [C] // Proceedings of the 6th International Conference on Neural Information Processing Systems. Denver: ACM, 1993: 737-744.
- 30 Jindal S, Manduchi R. Contrastive representation learning for gaze estimation [J]. *Proc Mach Learn Res*, 2023, 210: 37-49.
- 31 Liu S, Mallol-Ragolta A, Parada-Cabaleiro E, et al. Audio self-supervised learning: a survey [J]. *Patterns*, 2022, 3 (12): 100616.
- 32 Grill JB, Strub F, Alché F, et al. Bootstrap your own latent a new approach to self-supervised learning [C] // Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: ACM, 2020: 21271-21284.
- 33 Bardes A, Ponce J, LeCun Y. VICReg: variance-invariance-covariance regularization for self-supervised learning [EB/OL]. <http://arxiv.org/abs/2105.04906>.
- 34 Caron M, Misra I, Mairal J, et al. Unsupervised learning of visual features by contrasting cluster assignments [C] // Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: ACM, 2020: 9912-9924.
- 35 Jure Z, Li J, Ishan M, et al. Barlow twins: self-supervised learning via redundancy reduction [J/OL]. <https://doi.org/10.48550/arXiv.2103.03230>.
- 36 Barlow HB, Tolhurst DJ. Why do you have edge detectors? [C] // Optical Society of America Annual Meeting. Albuquerque: Optica Publishing Group, 1992.
- 37 Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex [J]. *J Physiol*, 1968, 195 (1): 215-243.
- 38 Marr D, Hildreth E. Theory of edge detection [J]. *Proc R Soc Lond B Biol Sci*, 1980, 207 (1167): 187-217.
- 39 Zylberberg J, Murphy JT, DeWeese MR. A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields [J]. *PLoS Comput Biol*, 2011, 7 (10): e1002250.
- 40 Mijatovic G, Antonacci Y, Loncar-Turukalo T, et al. An information-theoretic framework to measure the dynamic interaction between neural spike trains [J]. *IEEE Trans Biomed Eng*, 2021, 68 (12): 3471-3481.
- 41 Larsson DT, Maity D, Tsiotras P. A generalized information-theoretic framework for the emergence of hierarchical abstractions in resource-limited systems [J]. *Entropy*, 2022, 24 (6): 809.
- 42 Chen KX, Beyeler M, Krichmar JL. Cortical motion perception emerges from dimensionality reduction with evolved spike-timing-dependent plasticity rules [J]. *J Neurosci*, 2022, 42 (30): 5882-5898.
- 43 Frady EP, Kleyko D, Sommer FT. Variable binding for sparse distributed representations: theory and applications [J]. *IEEE Trans Neural Netw Learn Syst*, 2023, 34 (5): 2191-2204.
- 44 Lima B, Florentino MM, Fiorani M, et al. Cortical maps as a fundamental neural substrate for visual representation [J]. *Prog Neurobiol*, 2023, 224: 102424.
- 45 Cavanagh SE, Hunt LT, Kennerley SW. A diversity of intrinsic timescales underlie neural computations [J]. *Front Neural Circuits*, 2020, 14: 615626.
- 46 Shepherd GMG, Yamawaki N. Untangling the cortico-thalamo-cortical loop: cellular pieces of a knotty circuit puzzle [J]. *Nat Rev Neurosci*, 2021, 22 (7): 389-406.
- 47 Balas B, Saville A. Neural sensitivity to natural image statistics changes during middle childhood [J]. *Dev Psychobiol*, 2021, 63 (5): 1061-1070.

(责任编辑: 孙菲)