arXiv:2301.02196v1 [q-bio.NC] 5 Jan 2023

通过竞争学习为联想记忆生成稀疏表示

LuisSacouto,Andreas WichertINESC-ID & 计算机科学与工程系，高等技术学院，里斯本大学（IST-UL），葡萄牙里斯本 {luis.sa.couto, andreas.wichert}@tecnico.ulisboa.pt

摘要

赫布学习是最为确立的大脑原理之一，它引出了神经元集群这一理论概念。基于这一概念，许多有趣的大脑理论应运而生。Palm 的研究通过二元联想记忆实现了这一概念；这一模型不仅具有广泛的认知解释力，还能提出神经科学预测。然而，联想记忆只能处理对数稀疏表征，因此极难将该模型应用于真实数据。我们提出了一种生物学上合理的网络，可将图像编码为适用于联想记忆的编码。该网络由多个神经元群组构成，各群组专门处理局部感受野，并通过竞争机制进行学习。在两个视觉数据集上进行自联想和异联想实验后，我们得出结论：该网络不仅优于稀疏编码基线方法，其性能还接近使用最优随机编码所达到的水平。

关键词 稀疏编码 · 联想记忆 · 自编码器 · 受大脑启发的模型

1 引言

随着深度学习的兴起，人工智能系统在解决视觉模式识别等感知任务方面表现得越来越出色。尽管这些系统在许多任务中取得了巨大成功，但仍缺少生物大脑所具备的一些理想特性，其中包括行为灵活性和能效[1]。因此，计算神经科学成为基础研究的关键领域，这不仅是出于“理解大脑”的视角，也是为了尝试开发能够“运作”的智能系统。
这两种视角的交集并不像它们各自那样受到关注，部分原因在于，似乎没有任何技术能在最常见的机器学习基准测试中取得深度学习的成果；此外，大脑作为一种智能系统，至今仍是现代科学的一大谜团。

尽管这一事实令人沮丧，但关于大脑仍有一些依据充分且广为接受的发现。其中最重要的一项是赫布学习的观点，该观点最初由 Hebb 提出 [2], ，后来得到多项研究的支持 [3, 4, 5, 6, 7, 8, 9]。这一原理指出，当两个神经元同时放电时，它们之间的突触会增强，从而使二者的相关性提高。

基于这一思路，许多作者提出了细胞集群的概念：神经元的扩散激活作为概念的分布式表征，大脑通过这些表征与概念互动 [2, 10]。由此，Hofdstater 和 Palm 等人 [11, 12] 提出，这些集群构成了大脑的语言。

Palm 的研究更进一步，提出了一种非常完整的智能系统实现方案，该系统以神经元集群为基础运行。他的模型的核心是联想记忆 [13, 14, 15]。更具体地说，是 Willshaw联想记忆模型 [16]。我们将在下一节详细介绍；目前只需指出，这种非常简单的二元记忆在信息存储容量和实现所需时间方面都极其高效

且能耗低。这一简单组件构成了一个非常完整的认知模型，能够以符合生物学约束的方式解决智能任务，并对神经科学提出预测（下一节将进一步讨论）。
尽管联想记忆模型能力卓越，但它有一个显著的局限：只有在处理对数级稀疏数据时，才能达到最佳性能，而现实数据远没有那么稀疏（见第 2 节）。因此，围绕该模型的研究大多要么停留在理论层面，要么使用随机生成的稀疏数据。于是，生成合适表征的问题便交由假设中的编码器处理，这些编码器将扮演大脑皮层中感觉特异性区域的角色。这种抽象方式在脑模型中很常见，例如 Hawkins 提出的、颇具争议的层次时序记忆模型 [17, 18]。

能够生成这些极其稀疏表征的编码器，将释放联想记忆的全部潜能，并由此使 Palm 等人的模型达到前面提到的交汇点：既能帮助我们理解大脑，又能构建在真实数据上表现良好的系统。

那么，这样的编码器能实现吗？并不容易。要理解原因，不妨把编码看作一组神经元的激活向量。每个输入仅仅激活少量活跃神经元还不够。尽管这样的向量确实很稀疏，因为其中只有少量的 1，但如果所有可能的输入总是激活同一组神经元，那么实际使用的空间区域就很小，而那些 0 只是人为填充的部分。

因此，我们需要让每个神经元的放电概率都大致相同且非常低。只有这样，才能以稀疏但分布良好的方式利用整个空间。此外，编码还需要保留相似性。也就是说，如果两个输入在某种意义上相似，那么它们的编码之间应有大量比特重叠 [19, 20]。

所有这些约束使得这个问题既重要又复杂。尽管已有一些生成稀疏编码的技术，正如我们将要看到的，它们并不适用于联想记忆所要求的对数稀疏度约束。在本文中，我们将着手解决这一问题，并提出一种构建稀疏编码的新方法，以满足这些严苛的要求。

2 联想记忆与神经元集群背景

尽管联想记忆在当前研究中已不那么热门，但它们曾一度是神经计算的主要研究焦点[13, 14]。它们大多被搁置一旁，因为与多层感知机或卷积网络等更复杂的网络不同，它们无法在真实且具有挑战性的任务中取得良好表现。

联想记忆背后的理念，是从极其简单的架构中尽可能榨取性能，因此这些网络在设计上非常简单。这使它们极为高效，也能通过速度极快的硬件实现。
联想记忆的目标是将向量 $\mathbf{x} \in \mathbf{R}^N$ 与向量 $\mathbf{y} \in \mathbf{R}^M$关联起来。更具体地说，它应当存储一组 L 对 $\{(\mathbf{x}^{(l)}, \mathbf{y}^{(l)})\}_{l=1}^{L}$，这样在给出提示向量 $\mathbf{x}$ 时，就能输出与之对应的关联向量$\mathbf{y}$。

当两个向量不同时，这种记忆称为异联想；而当一个模式与自身关联时，则称为自联想。为什么要让一对模式与自身关联？这样一来，记忆便能根据只包含完整模式一小部分的提示向量，回忆出完整模式。我们稍后会看到，这种用法非常实用。

多年来，人们提出了许多联想记忆模型 [13, 14]。总体而言，它们都属于两类之一——前馈或循环。前者是更简单的机器，可一次性计算出输出。这类记忆的一个典型例子是 Willshaw 联想记忆 [16]。另一方面，后者是一种动力系统，会持续演化，直到收敛到稳定状态；Hopfield 网络就是一个典型例子。

乍看之下，这或许有悖直觉，但我们将看到，在某些条件下，Willshaw 联想记忆的信息容量远超大多数循环记忆 [14]。此外，它结构简单，在计算和能耗方面都极为高效；结合其他特性来看，它确实是一个值得深入研究的模型。鉴于它是本文的核心记忆，从这里开始，我们将简称其为联想记忆；每当提及其他类型的记忆时，都会明确说明。

2.1 联想记忆网络

联想记忆是一种具有二值权重的前馈网络。给定输入向量 $\mathbf{x} = [x_1, x_2, \cdots, x_N]^T$，其麦卡洛克-皮茨神经元 [21] 根据公式 1 计算输出，其中 H 是左连续零一阶跃函数，而 T_j 是神经元特定的阈值。

$$y_j = H\left(\sum_{i=1}^{N} w_{ij}x_i - T_j\right) = \begin{cases} 1 & \sum_{i=1}^{N} w_{ij}x_i \geq T_j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

作为一种学习机制，该记忆采用典型的赫布学习：如果在训练过程中，两个神经元同时放电 [2]，那么它们之间就会存在突触 x_i andy_j。更具体地说，给定一个训练对 (**x**，**y**)，一个通用权重 w_{ij} 将根据方程 2 进行更新。

$$w_{ij}^{new} = \begin{cases} 1 & w_{ij}^{old} = 1 \\ 1 & x_i y_j = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

研究表明，这个简单模型具有巨大的存储容量。这里我们将简要概述存储容量的推导，旨在让读者关注使该模型达到最佳性能的关键假设。

不失一般性，我们聚焦于这样一种情况：要建立关联的向量彼此不同，但大小相同N。每当我们向记忆呈现一对向量 (**x**，**y**) 时，实际上存储的是**y**，因为记忆需要将**x** 作为提示向量输入。因此，实际上记忆存储的是 L 个向量。假设所有向量恰好都具有 M 个激活位，那么一个向量的信息量等于 $\log_2 \binom{N}{M}$。因此，记忆中存储的信息量将等于 $I = L\log_2 \binom{N}{M}$。

现在，假设 M 比特可以以相同的概率出现在任意向量位置，也就是说，每个维度的使用相对频率约为 $\frac{M}{N}$。据此，当且仅当不存在任何一对满足 $x_i = 1$ 和 $\overline{y_j} = 1$的情况时，一般权重 w_{ij} 才等于零。 $x_i = 1$ 为 $\frac{M}{N}$ 的概率与 $y_i = 1$的概率相同。因此，假设二者相互独立，则突触处于开启状态的概率等于联合 $\frac{M^2}{N^2}$。所以，在这个特定示例中，权重处于关闭状态的概率为 $1 - \frac{M^2}{N^2}$。现在，若要使权重在整个学习过程结束后仍处于关闭状态，并假设样本相互独立且同分布，则 $P\left[w_{ij} = 0\right] = \left(1 - \frac{M^2}{N^2}\right)^L$。因此，

$$p = P\left[w_{ij} = 1\right] = 1 - \left(1 - \frac{M^2}{N^2}\right)^L.$$

在回忆时，当且仅当满足以下条件时，记忆会在输出中产生一个伪位 $y_k = 1$： $\forall_{i=1}^{N} x_i = 1$： $w_{ik} = 1$。该事件的概率为 p^M。

将可接受伪位的数量上限设为 1（即假设 $(N - M)p^M = 1$），便可得出最重要几项指标的具体数值。例如，每个模式的激活位最优数量为 $M = \log_2 N - 2$。这说明向量必须是对数稀疏的。此外，可存储模式的最大数量为 $L = \log_2 \left(\frac{N^2}{M^2}\right)$。最后，在这些最优假设下，信息容量为 $I = N^2 \log_2 2,$，对于一个 $N \times N$ 位矩阵，其每比特容量为 $\log_2 \approx 69.31\%$。这个数值远高于大多数联想记忆；考虑到该模型的简单性及其生物学合理性，这一点尤其重要。总而言之，该模型仅采用成熟的赫布学习，可在神经硬件上实现，而且与大脑一样，具有极高的能效 [12]。

2.2 通过联想记忆形成神经元集群

许多研究在大脑中发现了提取特定特征的神经元 [12]。然而，没有研究发现只编码某个特定概念的神经元。此外，即使存在这样的神经元，鉴于神经元每天都会死亡，大脑也不会非常稳健，因此我们人类可能会不断遗忘概念。相比之下，通过分布式表征来编码信息，是一种更加合理、稳健且高效的方式 [19, 18]。

将这一事实与如今普遍认可的大脑中存在赫布学习这一观点结合起来 [4, 8, 9], , 许多作者提出了神经元集群的概念 [10, 22, 23, 24, 25]。这一概念指的是一组神经元，它们通过学习形成了紧密连接，以至于激活其中一部分就会导致整个群体被激活。
作为概念的分布式表征，这些细胞集合由单个神经元构成，而这些神经元也可以参与许多其他细胞集合。

这一看似简单的概念可以成为连接实验神经科学层面与信息处理层面的桥梁 [11, 12]。与单个神经元相比，细胞集合的粒度更粗，更适合用于对认知和思维过程进行更高层次的推理。

由于联想记忆以赫布学习和分布式表征为基础，因此它是模拟细胞集合的理想选择，而 Palm 的研究工作一直聚焦于此 [10, 25]。具体来说，一个细胞集合的点燃可以视为自联想的实例，而这些细胞集合的序列则对应于异联想。加入在生物学上合理的阈值控制机制后 [25], , Palm 的团队构建了具有很强解释能力的模型，将大脑皮层视为由联想记忆组成的群体。这类思想已应用于语言理解等领域 [26, 27]。此外，联想记忆模型的本质也促成了许多得到广泛证实的神经科学预测。其中有两点值得强调：第一，大多数任务都必须涉及许多大脑区域[28, 29]；第二，正如我们在上一小节中看到的，大脑中的分布式表征应当是稀疏的 [30, 31, 32, 33, 34, 35]。

稀疏表示预测结果得到验证，这让我们更加确信联想记忆是大脑皮层的一个基础构件。然而，正如我们之前所说，这种稀疏编码要求 [36, 37, 20, 38]也是其在实际应用中的主要障碍。在下一节中，我们将尝试解决这个问题。

3 所提出的网络

一组神经元可以通过学习一组最能表征向量$\mathbf{x} \in \mathbb{R}^N$的原型 $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N\}$ 来表征这些向量。学会一个字母表后，编码就是一个二元存在向量$\mathbf{z}$，它根据某个距离函数 $D(\mathbf{w}_i, \mathbf{x})$，指示当前输入与哪个原型最相似。具体而言，编码向量 z_i 的每个元素都是方程 3 中的指示函数，这类神经元组通常称为胜者全得竞争层 [14, 39]。

$$z_i = \mathbb{I}\left(i = \underset{1 \le j \le N}{\arg\min}\, D\left(\mathbf{w}_j, \mathbf{x}\right)\right) \tag{3}$$

竞争层已得到广泛研究，其学习过程可以通过侧向抑制 [14]以符合生物学合理性的方式实现。概括而言，这个过程相当简单。给定一个输入，各神经元会竞争着对其进行表征；按照最近邻的意义，最适合的神经元获胜并进行更新。其他神经元则保持不变。更新规则是方程 4 中简单的迭代均值，它使获胜神经元更适合表征当前输入。

$$\mathbf{w}_i^{new} = \mathbf{w}_i^{old} + \alpha\left(\mathbf{x} - \mathbf{w}_i^{old}\right) z_i \tag{4}$$

采用“赢家通吃”方法时，编码是一个指示获胜者的独热向量。因此，这些编码是稀疏的，即只有一个有效位，其余大多为零。然而，没有任何机制能保证神经元以相同的频率激活。具体来说，在对输入向量 $\mathbf{x}$ 进行采样时，我们希望权重向量 $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N\}$ 满足这样的条件：每个神经元赢得竞争的概率都是 $\frac{1}{N}$。幸运的是，DeSieno [40], 通过惩罚过于频繁激活的神经元，解决了这一等概率问题。
这种活动依赖偏置在生物学上也具有合理性，并且可以通过局部学习规则实现。具体而言，我们使用公式 5 来持续估算每个神经元的活动程度。随后利用这一平均活动 f_i，根据公式 6 计算偏置 b_i（其中 γ 控制偏置在竞争中的权重）。加入偏置后，获胜神经元不再是使距离 D最小的神经元，而是使 $D(\mathbf{w}_i, \mathbf{x}) - b_i$最小的神经元。

$$f_i^{new} = f_i^{old} + \beta\left(z_i - f_i^{old}\right) \tag{5}$$

$$b_i = \gamma\left(\frac{1}{N} - f_i\right) \tag{6}$$

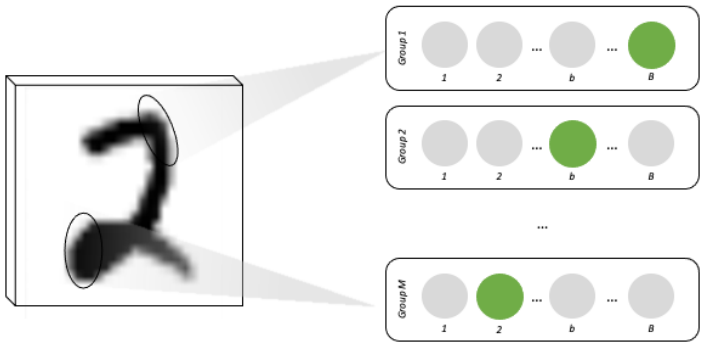

图1：图像由一组 M 神经元组进行编码。每个神经元组聚焦于一个感受野，并试图表示其中的内容。为此，在每个神经元组内，神经元相互竞争，以表示当前输入。胜者通吃，单独放电。最终，所有激活构成的向量将成为二进制稀疏编码，其中每个神经元的放电频率大致相同。

加入这种偏置后，胜者全得竞争层可以生成这样的编码：神经元以稀疏且频率非常相近的方式放电。那么，我们达成目标了吗？还没有，因为相似性保持这一关键属性仍然缺失。独热编码是其输入的准符号表示，因为它们之间不存在部分相似性。具体来说，当两个输入之间的距离非常小时，它们会产生同一个获胜神经元，因此得到相同的编码。在这种情况下，编码会发生重叠，产生 1 比特的相似度分数。

然而，当两个输入的差异达到某个程度，以至于二者永远不会发生重叠时，无论它们实际上有多相似，相似性始终都会是零。考虑到这一点，可提取的原型数量可能会受到严重限制，进而限制我们能够建模的领域的丰富程度。当原型过多时，其中一些必然会彼此相似。在这种情况下，两个非常相似的输入生成两种不同编码的概率会很高。

此时，我们已经能够学习神经元群组，将某些输入编码为稀疏且频率均匀的放电编码。然而，我们还没有办法使这些编码保留相似性。现在，这种方法将生成仅包含一个比特的编码。如果不使用一个神经元组对输入进行编码，而是使用 M，那么从激活位所占百分比来看，我们的向量仍然同样稀疏，而且各个位的使用也同样均匀。我们可以利用这一原则向编码中添加激活位，从而实现更丰富的描述。

然而，如果所有神经元组的输入都相同，那么它们的学习结果将几乎完全相同，使用 M 比特而非一个比特也就多此一举。为了解决这个问题，我们借鉴生物学的一项原理，让每个神经元组专注于一个随机分配的局部感受野。这样，每个组就会对视野中的一个区域进行建模，从而将问题拆解为复杂度更低的子问题。在最终编码中，每组的获胜者将代表一个视觉区域的内容，因此可作为已学习特征的二进制指示符。

图1展示了所提出编码器网络的示意图。在此示例中，我们描绘了三个神经元组，每组包含 B 个元素。若要构建一个 H 维编码器，例如可以选择 B，使其近似等于 $\frac{H}{M}$。图中还展示了第一组和最后一组的感受野。每组都关注图像中的一个区域，并利用其中的获胜神经元对该区域的内容进行编码。通过这种安排，稀疏编码能够表示不同输入之间更多层次的相似性，从而解决我们的问题。

在详细阐述了所提编码器的数学原理后，下一节我们将转向实证评估。我们将考察直觉是否成立、所提模型与同类方法相比表现如何，以及它是否最终能让联想记忆应用于真实数据成为可能。

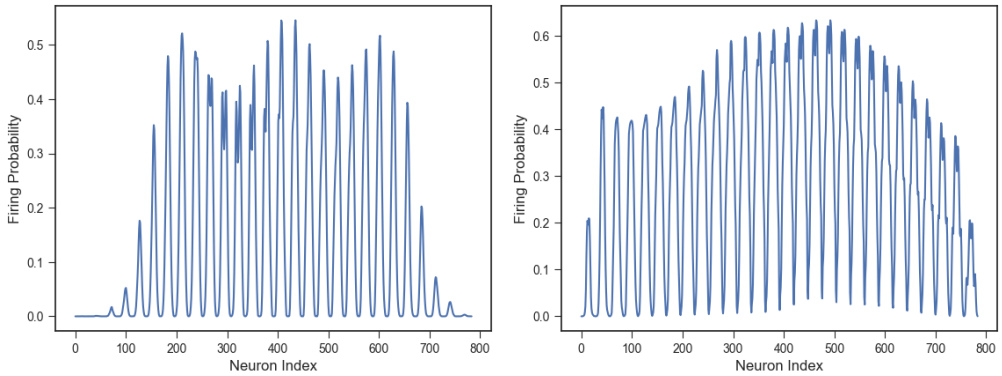

图 2：假设每个神经元编码一个像素，我们可以绘制它们在常规图像中的使用情况。这样一来，我们可以看到，对于真实数据，各神经元的放电概率既不低，也不相近。左侧展示的是 MNIST 数据集上的情况，右侧则展示 Fashion-MNIST 上的相同结果。

4 实验分析

在本节中，我们将使用两个广为人知的视觉模式识别数据集，探索所提方案的能力：MNIST [41] 和 Fashion-MNIST [42]。两个数据集均由 70000 张灰度图像组成，每张图像包含 28×28 个像素。假设每个像素对应一个神经元，我们选取这些样本，并在图 2 中绘制了两个数据集的相对放电频率。通过分析结果，我们可以清楚地看到，为什么我们会说真实数据既不稀疏，也不会在各个向量维度上均匀分布。

4.1 与深度学习基线的比较

为了恰当地评估所提出的联想记忆编码的质量，我们需要一个基线。为此，我们选择了深度自编码器 [43]，它们不仅最小化重建损失，还最小化平均放电频率惩罚项；该惩罚项采用公式 7 中所示的批次级 KL 散度形式（其中 ρ 是期望的放电相对频率，L 是批次大小，H 是编码层的大小）。此外，为了确保编码为二进制，我们在中间隐藏层 [44] 使用 Gumbel Softmax 激活函数。

$$\sum_{l=1}^{L}\sum_{i=1}^{H}\left(\rho * log\left(\frac{\rho}{h_{li}}\right)+(1-\rho)*\log\left(\frac{1-\rho}{1-h_{li}}\right)\right) \tag{7}$$

确立基线架构后，还有许多可以调整的超参数。编码层 H的大小随机选取以下三种之一：512、1024 和 2048。由此，我们可以计算出所需的活跃神经元数量，约为 $M=\log_2\binom{H}{4}$（见第 2 节）。确定 M后，我们可以定义目标放电相对频率 $\rho=\frac{M}{H}$。层数为 2 到 8 之间的随机偶数，每层的神经元数量从输入大小到 H线性变化。训练时，我们还会对以下参数进行不同的采样：训练轮数、学习率、重建损失、批量大小以及稀疏性约束的相对强度[1]。

图3中的每个绿点都对应一个经过上述采样步骤后完成训练的自编码器。蓝点表示我们提出的方法的运行结果，其中所有超参数也都是从合理取值区间内随机采样的。最后，红点对应于随机生成且完全满足联想记忆要求的数据。分析该图可见，我们的方法不仅明显优于自编码器基线，而且结果也非常接近最优上限。

如果我们从三种编码中各选取表现最佳的运行结果（确保它们的维度相同），就可以进行图 4 所示的更详细比较。观察每个神经元的放电频率，我们便能看出所提方案有多接近最优。此外，我们可以看到，它能更好地保留编码空间中图像之间的相似性。

[1]所有模拟的代码　　可通过发送电子邮件向第一作　　作者

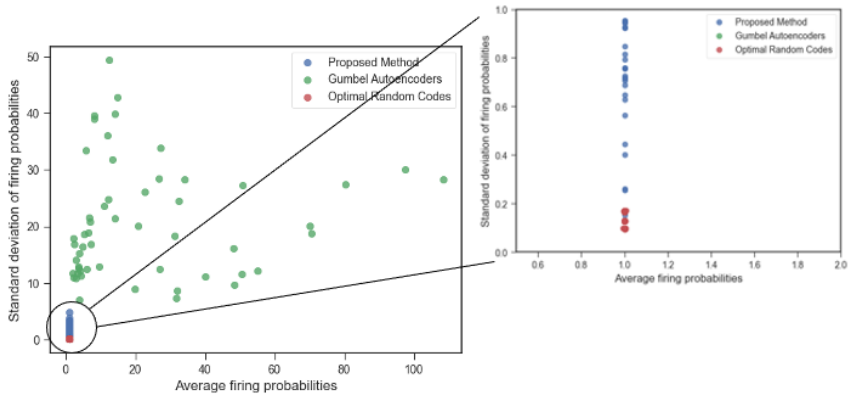

图3：在 MNIST数据集上，三个模型在多次运行中的神经元放电概率的均值和标准差：Gumbel自编码器基线、所提方法和最优随机编码。

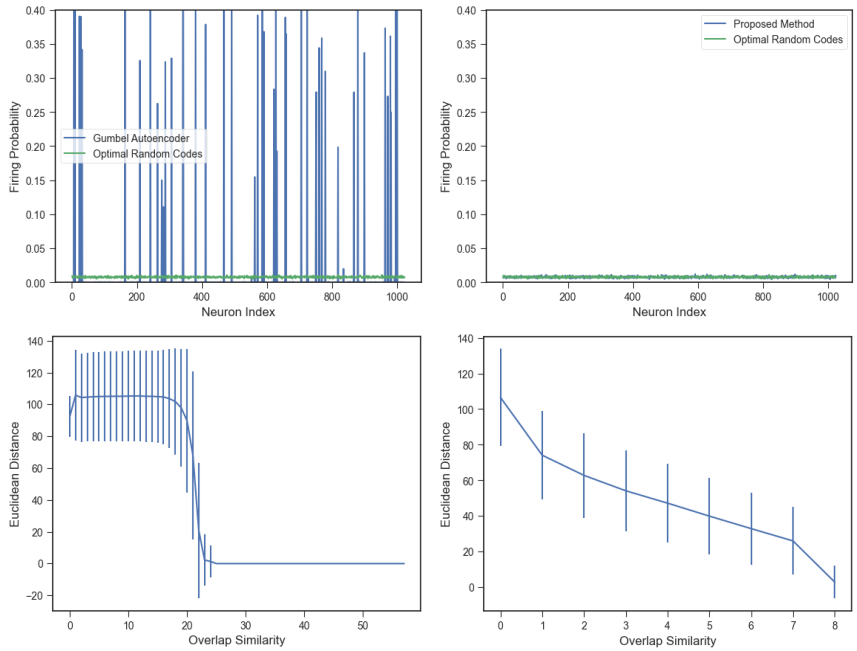

图 4：顶行展示了最佳自编码器（左）和所提方案最佳运行结果（右）中各个神经元的放电概率。此外，还将两者与最优随机编码的概率进行了比较。所提方法更接近目标，几条线几乎完全重叠。底行展示了图像空间中的欧几里得距离与编码之间比特重叠的关系。尽管两种方法都保留了一定的相似性，但所提方法（右）的距离会随着相似性的增加而逐渐减小，而自编码器（左）则表现出近乎二元的行为。

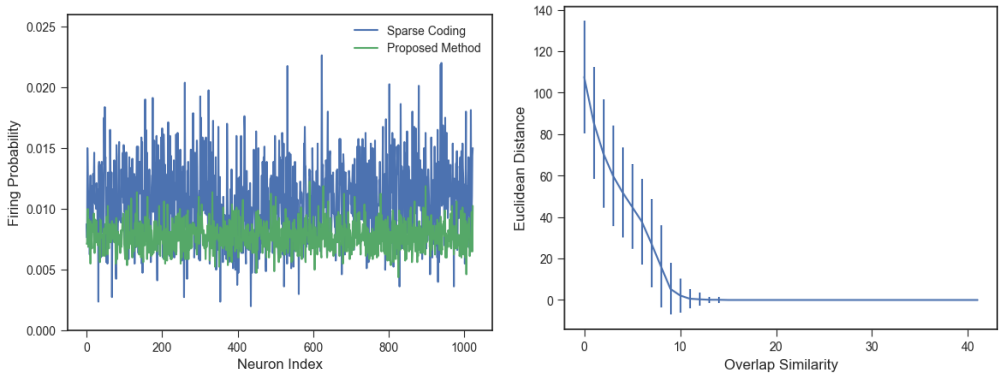

图 5：左图比较了所提出的方法与基于坐标下降的稀疏编码中各个神经元的放电概率。右图展示了经典稀疏编码中输入空间距离与重叠相似性之间的关系。

4.2 与经典稀疏编码的比较

另一种非常常见的稀疏编码方法是字典学习 [45]。在这种设定下，我们学习由特征组成的字典 $\mathbf{V}$，并学习尽可能好的稀疏编码 $\mathbf{U}$，使得 $\mathbf{UV}$ 能够近似输入 $\mathbf{X}$。这对应于方程 8 中写出的损失函数。$\mathbf{U}$ 和 $\mathbf{V}$ 都必须通过优化求得，通常采用交替坐标下降法。

$$\frac{1}{2}\|\mathbf{X}-\mathbf{UV}\|_F^2+\alpha\|\mathbf{U}\|_1 \tag{8}$$

通过在超参数空间中进行随机搜索，我们找到了这种经典稀疏编码技术的一个合适版本，将其用作与我们的方法进行比较的基准。为确保比较公平，我们确保所有编码的维度都相同，均为 1024 个神经元。

4.2.1 神经元的放电频率

再次比较两种编码的神经元使用情况。分析图5可见，两种方法之间的差异远小于我们的方法与自编码器基线之间的差异。此外，我们还发现，经典稀疏编码也能够保留相似性。不过，正如接下来将看到的，在处理联想记忆时，这些看似微小的差异会带来巨大的代价。

4.2.2 在自联想中存储

现在，我们确认这两种技术都大致满足记忆的要求，可以将两组编码存储到记忆中。我们先从自联想开始，其中每个模式都与自身相关联，然后逐步删除比特，观察记忆能否补全缺失的信息。乍看之下，有两个指标似乎很重要：比特召回率和比特精度。然而，经过更详细的分析，我们发现记忆不会丢失作为提示向量提供的比特，因此召回率始终为100%。因此，在图6中，我们仅展示两种技术在30次运行中的平均比特精度，以及随着删除比特数量增加而产生的变化。无论记忆的填充程度如何，结果都明显有利于所提出的方法。此外，即使记忆中存储的模式数量约为神经元数量的8倍（本例中为1024），精度仍然很高。

自联想的一个典型用途是模式补全或插值。为便于说明，我们给出了图7，其中每个像素都有 50% 的概率被置为零。接着，我们对带噪模式进行编码，让记忆对其进行自联想，然后根据编码重建图像。结果非常理想；总的来说，比特精度比目视检查重建质量更严格。

4.2.3 在异联想中存储

所有自联想实验都很重要，但联想记忆最难完成的任务是异联想。采用经典稀疏编码技术进行的实验表明，该技术完全不适用于此目的

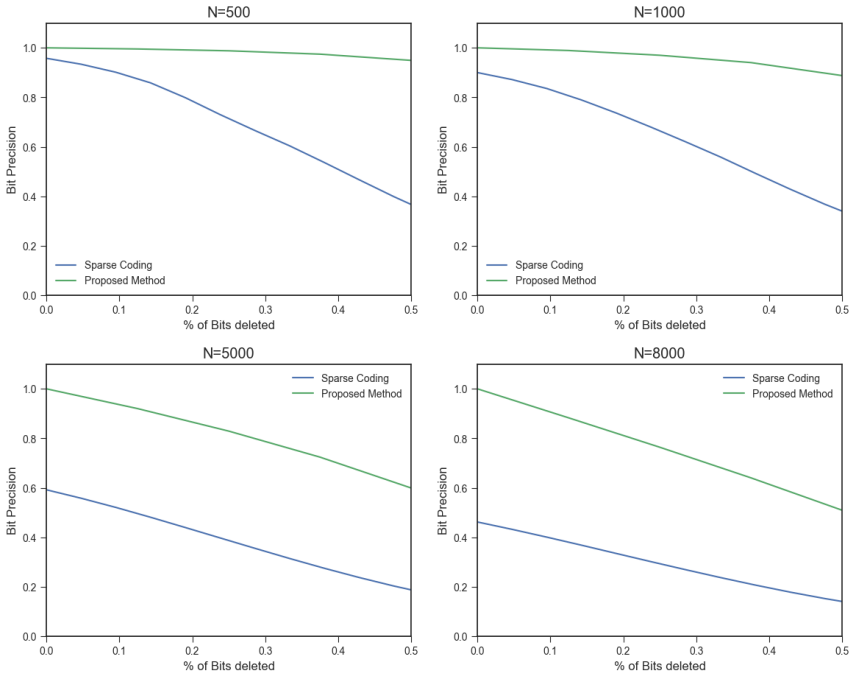

图6：随着已 存储示例数量增加，所提出的编码仍能保持较高的比特精度，而 s
经典稀疏码 会导致精度迅速下降。

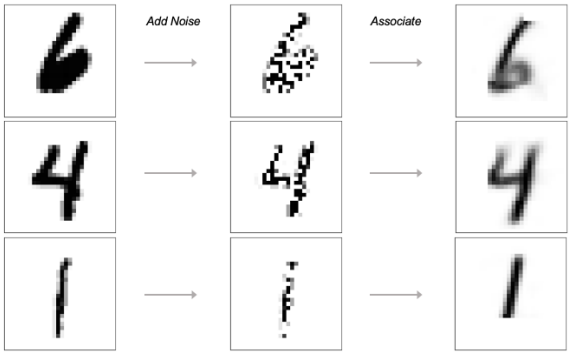

图7：一 使用我们的编码和自联想记忆进行模式补全的几个示例 n.

。因此，我们决定将我们的方法与最优随机编码进行比较。图8展示了异联想流程。为了将一对图像存入记忆，我们先对它们进行编码，然后以关联对的形式将相应的稀疏编码呈现给记忆。接着，我们用这一对中的一个成员提示记忆，并观察记忆能否回忆出正确的关联对象，以此衡量异联想质量。图中所示的例子都是真实的，且是通过我们提出的方法实现的。

与自联想不同，由于我们并非将一个模式与其自身关联，因此比特召回率无法保证完美。此外，衡量记忆的关联准确率，并了解它有多少次回忆出完全错误的模式，也很有意义。我们通过找出与记忆输出重叠最多的编码来定义这一指标。如果该编码有误，我们就计为一次错误。在图9中，我们展示了随着存储模式数量增加而变化的三项关键指标。与最优随机编码相比，我们的方法表现极为出色，尤其是在比特召回率和关联准确率方面。由于使用频率并非完全均匀，会产生伪相关，导致记忆极度饱和时一些错误的神经元被激活。

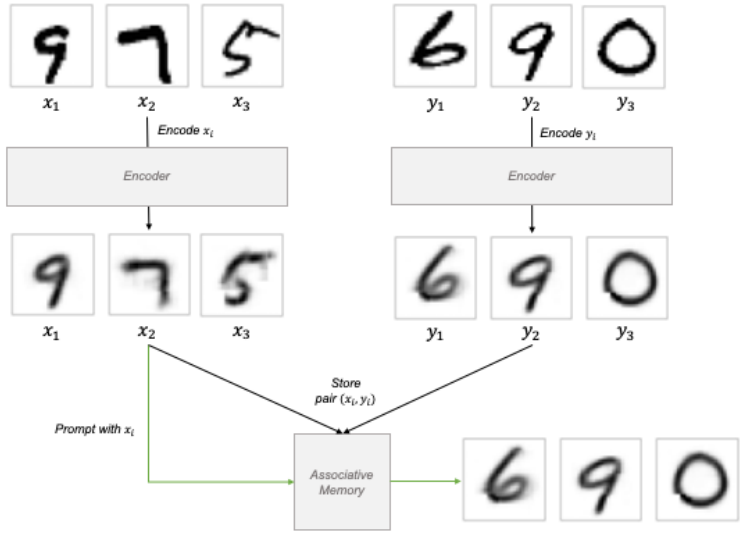

图8：每对待关联的图像都会先进行编码，然后呈现给记忆。记忆训练完成后，我们可以向其呈现一张提示图像，它会尝试重建与之配对的图像。图中展示了一些真实实验示例。

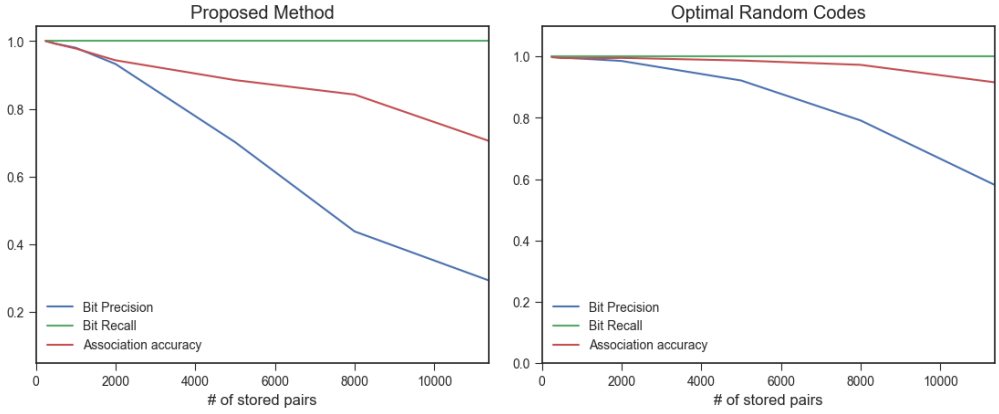

图9：左图跟踪使用所提出的编码训练的异质联想记忆的比特召回率、比特精度和关联准确率。右图跟踪使用最优随机编码时的相同指标。尽管所提出的编码并非最优，但它们使记忆表现得极具竞争力。

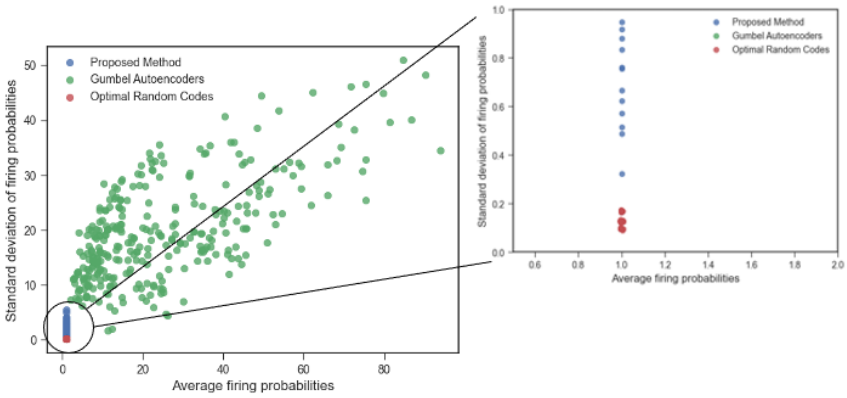

图10：在 Fashion-MNIST 数据集上，三个模型（Gumbel自编码器基线、本文提出的方法和最优随机编码）多次运行时神经元放电概率的均值和标准差。

4.3 扩展到更复杂的数据集

为了确保我们的分析有效且具有普适性，我们决定在一个更复杂的数据集——Fashion-MNIST 上开展非常相似的实验。

图10和图11从神经元相对放电频率和相似性保持两方面展示了与自编码器基线的比较。结论与 MNIST 的情况完全相同，因为这些图看起来几乎就是图3和图4的翻版。

图 12 展示了自联想中的记忆比特精度，即使在极端严苛的条件下——记忆极其饱和，且提示向量中的大量比特已被删除——仍然非常高。
最后，图13提供了与我们针对 MNIST 所做的相同异联想分析（见图9）。再次证明，我们的方法不仅能够实现大多正确的关联，在关联准确率方面也更具竞争力。这种现象很可能是因为，与较为均一的 MNIST 相比，该数据集的多样性更高。

5 结论

尽管人工智能不断进步，但没有任何系统能够达到人脑的灵活性和适应能力。因此，计算神经科学是一项关键研究领域，它不仅有助于揭开其中的奥秘，也能为开发更新、更出色的智能系统提供灵感。然而，大多数受生物学启发的系统在真实数据上的表现难以与深度学习等技术竞争。只有弥合这一差距，我们才能推动该领域迈上新的台阶。

Hebb 最早提出的神经细胞集群概念，催生了若干潜力巨大的脑模型。例如，Palm 的研究通过二元联想记忆实现了这一概念。这个开创性的模型提出了神经科学预测，并在解释思维过程的运作机制和认知能力方面展现出强大的解释力。

然而，这里有个问题。联想记忆只有在处理对数稀疏、频率均等且保持相似性的编码时才能发挥良好效果。鉴于真实数据远不具备这些特性，也不存在能够以这种方式对其进行充分映射的编码器，因此大多数相关研究都使用随机生成的数据。

我们的目标是构建一种编码器，能够接收真实的视觉数据，并将其表示为满足记忆严格约束的稀疏编码。具体来说，编码中的所有维度都应以大致相等的概率被使用，而且输入空间中的相似性应在编码空间中得到保留。

我们将一组神经元定义为一个“赢家通吃”的竞争层，其中通过 Desieno的思想——活动依赖偏置——确保放电概率相等。由此，我们注意到，如果将活跃/非活跃神经元视为比特，一组神经元就会产生稀疏且频率均等的编码。

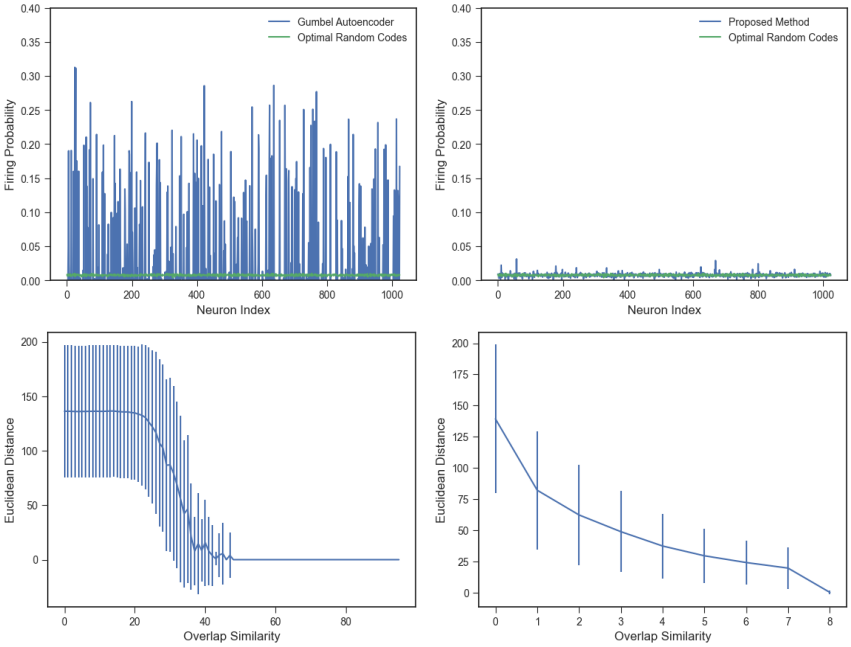

图 11：Fashion-MNIST 数据集上的图 4 复现。结论相同：所提出的方法明显优于基线。

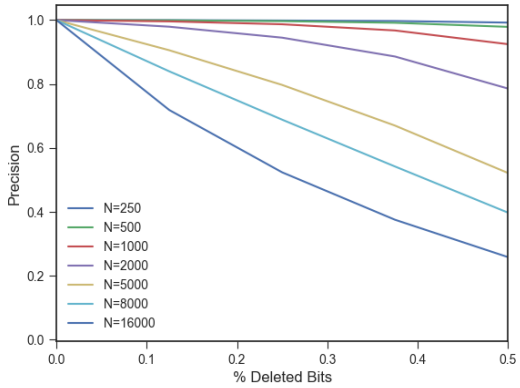

图 12：即使存储模式的数量远远超过单元数量（本例中为 1024），所提出的编码方法仍能保持较高的比特精度。

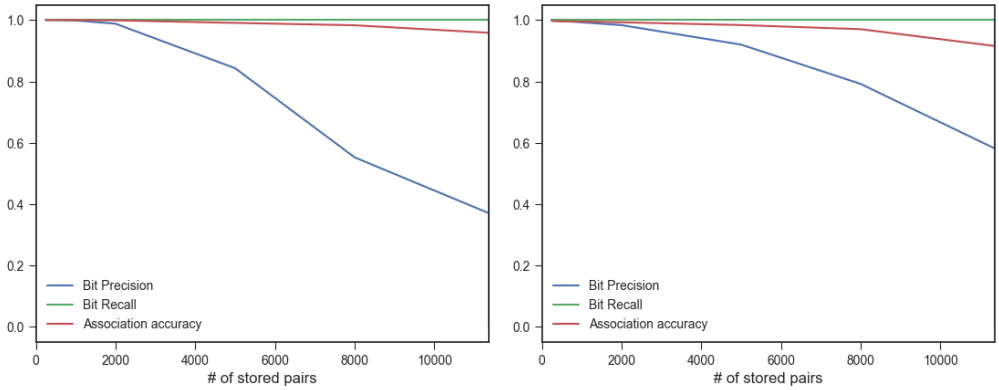

图13：我们再次使用三项关键性能指标，比较采用所提出的编码和最优随机编码时联想记忆的性能。不过，这次我们在 Fashion-MNIST 数据集上进行比较。

随后我们发现，要表示一个复杂领域，需要一个非常庞大的群组。随着群组规模增大，特征会变得更加具体，但独热编码无法表示局部相似性。为了解决这个问题，我们借鉴了生物学中的感受野概念，将神经元分成若干群组，每个群组专门负责视野中的特定区域。将这些群组组合起来后，我们构建了一个符合生物学特征的分布式网络，能够将图像编码为适用于联想记忆的稀疏编码。

为了验证我们的直觉，我们选取了两个知名数据集 MNIST 和 Fashion-MNIST，并将我们的方法生成的编码与两种替代方法进行了比较。与所提出的方法相比，坐标下降稀疏编码和 Gumbel Softmax 自编码器在稀疏性、放电概率的均匀性以及相似性保持方面表现更差。此外，所提出的方法与最优随机编码极为接近。获得这些编码后，我们继续测试它们在联想记忆中的表现。无论是在自联想还是异联想任务中，我们不仅可以说所提出的方法胜过了基线方法，而且其表现也接近最优。

总之，通过实现提出的目标，我们的方法为进一步探索采用细胞集合和联想记忆的脑模型打开了大门。未来的工作可以沿着这一方向继续探索，或通过以下两种方式进一步发展编码器。首先，我们可以组合神经元群组来编码更丰富的特征，并将其推广到更复杂的任务。其次，编码器可以应用于语音等其他类型的感知数据，无需进行重大改动。

致谢

第一作者将这项工作献给 Margarida；没有她，这项工作便不可能完成。此外，我们谨感谢葡萄牙科学技术基金会（FCT）对本项目的支持：该基金会向第一作者提供了博士生资助 SFRH/BD/144560/2019，并提供了常规资助 UIDB/50021/2020。该基金会未参与研究设计、数据收集与分析、发表决定或稿件撰写。作者声明不存在利益冲突。所有实验的代码和数据可通过电子邮件向第一作者索取。

参考文献

[1] Anthony Zador、Blake Richards、Bence Ölveczky、Sean Escola、Yoshua Bengio、Kwabena Boahen、Matthew Botvinick、Dmitri Chklovskii、Anne Churchland、Claudia Clopath、James DiCarlo、Surya Ganguli、Jeff Hawkins、Konrad Koerding、Alexei Koulakov、Yann LeCun、Timothy Lillicrap、Adam Marblestone、Bruno Olshausen、Alexandre Pouget、Cristina Savin、Terrence Sejnowski、Eero Simoncelli、Sara Solla、David Sussillo、Andreas S. Tolias和Doris Tsao。《迈向新一代人工智能：推动神经人工智能革命》，2022年10月。arXiv:2210.08340 [计算机科学，定量生物学].[2] D.O. Hebb.《行为的组织》。纽约，John Wiley & Sons，1949年。[3] Greg J. Stuart和Bert Sakmann。躯体动作电位向新皮质锥体细胞树突的主动传播。《自然》，367(6458)：69–72，1994年。出版方：自然出版集团。[4] Ami Citri和Robert C. Malenka。突触可塑性：多种形式、功能与机制。《神经精神药理学》，33(1)：18–41，2008年。出版方：自然出版集团。

[5] Randall C. O'Reilly。皮层认知生物学计算模型的六项原则。《认知科学趋势》，2(11)：455–462，1998年。出版方：爱思唯尔。[6] Botond Szatmáry和Eugene M. Izhikevich。工作记忆的脉冲时序理论。PLoS计算生物学，6(8)：e1000879，2010年。出版方：公共科学图书馆，旧金山，美国。[7] Omri Barak和Misha Tsodyks。工作记忆的工作模型。《神经生物学当前观点》，25：20–24，2014年。出版方：爱思唯尔。[8] Yuanyuan Mi、Mikhail Katkov和Misha Tsodyks。工作记忆容量的突触相关因素。《神经元》，93(2)：323–330，2017年。出版方：爱思唯尔。[9] Florian Fiebig和Anders Lansner。基于赫布式短时程增强的脉冲式工作记忆模型。《神经科学杂志》，37(1)：83–96，2017年。出版方：神经科学学会。[10] Günther Palm。细胞集合为脑研究提供的指引。《神经科学概念》，1：133–147，1990年。[11] Douglas R. Hofstadter。《哥德尔、艾舍尔、巴赫：集异璧之大成》。基础图书出版社，纽约，20周年纪念版，1999年。[12] Günther Palm。神经联想记忆：经典人工智能的另一种方法。施普林格·自然，2022年。[13] Günther Palm。论联想记忆。《生物控制论》，36(1)：19–31，1980年。出版方：施普林格。[14] Robert Hecht-Nielsen。《神经计算》。艾迪生-韦斯利朗文出版公司，1989年。[15] G. Palm。突触修饰局部规则的记忆容量。《神经科学概念杂志》，2(1)：97–128，1991年。[16] David J. Willshaw、O. Peter Buneman和Hugh Christopher Longuet-Higgins。非全息联想记忆。《自然》，222(5197)：960–962，1969年。出版方：自然出版集团。[17] Yuwei Cui、Subutai Ahmad和Jeff Hawkins。基于无监督神经网络模型的连续在线序列学习。《神经计算》，28(11)：2474–2504，2016年11月。arXiv:1512.05463 [计算机科学，定量生物学].[18] Subutai Ahmad和Luiz Scheinkman。我们为何如此稠密？使用高度稀疏表示的益处，2019年4月。arXiv:1903.11257 [计算机科学，统计学].[19] J. Hawkins、S. Ahmad、S. Purdy和A. Lavin。生物与机器智能（BAMI）。首次在线发布0.4版，2016年。[20] Günther Palm。神经联想记忆与稀疏编码。《神经网络》，37：165–171，2013年。出版方：爱思唯尔。[21] Warren S. McCulloch和Walter Pitts。神经活动内在观念的逻辑演算。《数学生物物理学通报》，5(4)：115–133，1943年。出版方：施普林格。[22] George L. Gerstein、Purvis Bedenbaugh和Ad MHJ Aertsen。神经元集合。《IEEE生物医学工程汇刊》，36(1)：4–14，1989年。出版方：IEEE。[23] Wolf Singer、Andreas K. Engel、Andreas K. Kreiter、Matthias HJ Munk、Sergio Neuenschwander和Pieter R. Roelfsema。神经元集合：必要性、特征标志与可检测性。《认知科学趋势》，1(7)：252–261，1997年。出版方：爱思唯尔。[24] Damian J. Wallace和Jason ND Kerr。追踪细胞集合。《神经生物学当前观点》，20(3)：296–305，2010年。出版方：爱思唯尔。[25] Günther Palm、Andreas Knoblauch、Florian Hauser和Almut Schüz。大脑皮层中的细胞集合。《生物控制论》，108(5)：559–572，2014年。出版方：施普林格。[26] Rebecca Fay、Ulrich Kaufmann、Andreas Knoblauch、Heiner Markert和Günther Palm。在神经机器人系统中结合视觉注意、物体识别与联想信息处理。载于David Hutchison、Takeo Kanade、Josef Kittler、Jon M. Kleinberg、Friedemann Mattern、John C. Mitchell、Moni Naor、Oscar Nierstrasz、C. Pandu Rangan、Bernhard Steffen、Madhu Sudan、Demetri Terzopoulos、Dough Tygar、Moshe Y. Vardi、Gerhard Weikum、Stefan Wermter、Günther Palm和Mark Elshaw编，《智能机器人的仿生神经学习》，第3575卷，第118–143页。施普林格·柏林·海德堡，柏林，海德堡，2005年。丛书：《计算机科学讲义》。[27] Andreas Knoblauch、Heiner Markert和Guenther Palm。皮层语言与动作处理的联想模型。载于《语言、认知与动作建模》，第79–83页，英国普利茅斯大学，2005年5月。世界科学出版社。

[28] Amos Arieli、Alexander Sterkin、Amiram Grinvald 和 A. D. Aertsen。持续活动的动态：解释诱发皮层反应的巨大变异性。《科学》，273(5283):1868–1871，1996。出版方：美国科学促进会。[29] Henry Markram。蓝脑计划。《自然综述：神经科学》，7(2):153–160，2006。出版方：自然出版集团。[30] Bruno A. Olshausen 和 David J. Field。通过学习自然图像的稀疏编码，涌现简单细胞的感受野特性。《自然》，381(6583):607–609，1996。出版方：自然出版集团。[31] Benjamin Willmore 和 David J. Tolhurst。神经编码稀疏性的表征。《网络：神经系统计算》，12(3):255，2001。出版方：IOP 出版社。[32] Stephen Waydo、Alexander Kraskov、Rodrigo Quian Quiroga、Itzhak Fried 和 Christof Koch。人类内侧颞叶中的稀疏表征。《神经科学杂志》，26(40):10232–10234，2006。出版方：神经科学学会。[33] R. Quian Quiroga、Leila Reddy、Gabriel Kreiman、Christof Koch 和 Itzhak Fried。人脑中单个神经元的不变视觉表征。《自然》，435(7045):1102–1107，2005。出版方：自然出版集团。[34] R. Quian Quiroga、Gabriel Kreiman、Christof Koch 和 Itzhak Fried。内侧颞叶中的稀疏编码，但并非“祖母细胞”编码。《认知科学趋势》，12(3):87–91，2008。出版方：爱思唯尔。[35] Rodrigo Quian Quiroga。概念细胞：陈述性记忆功能的构成基石。《自然综述：神经科学》，13(8):587–597，2012。出版方：自然出版集团。[36] Gunther Palm。利用神经网络进行计算。《科学》，235(4793):1227–1228，1987。出版方：美国科学促进会。[37] Günther Palm、Friedhelm Schwenker 和 Friedrich T. Sommer。联想记忆网络与稀疏相似性保持编码。载于《从统计学到神经网络》，第282–302页。施普林格，1994。[38] Luis Sa-Couto 和 Andreas Wichert。在 Willshaw 联想网络中存储对象依赖型稀疏编码。《神经计算》，32(1):136–152，2020年1月。[39] Luis Sa-Couto 和 Andreas Wichert。注意力启发网络：不变模式识别模型的陡峭学习曲线。《神经网络》，114:38–46，2019年6月。[40] DeSieno。为竞争学习添加良知机制。载于*IEEE* 神经网络国际会议，第1卷，第117–124页，美国加利福尼亚州圣迭戈，1988年。IEEE。[41] Yann LeCun、Corinna Cortes 和 Chris Burges。MNIST 手写数字数据库。[42] Han Xiao、Kashif Rasul 和 Roland Vollgraf。Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集，2017。[43] Ian Goodfellow、Yoshua Bengio 和 Aaron Courville。深度学习。麻省理工学院出版社，2016。[44] Eric Jang、Shixiang Gu 和 Ben Poole。基于 Gumbel-Softmax 的分类重参数化。*arXiv* 预印本 *arXiv:1611.01144*，2016。[45] Julien Mairal、Francis Bach、Jean Ponce 和 Guillermo Sapiro。用于稀疏编码的在线字典学习。载于第*26*届年度国际机器学习会议论文集 - *ICML ’09*，第1–8页，加拿大魁北克省蒙特利尔，2009。ACM 出版社。