arXiv:2301.08530v2 [physics.data-an] 1 Apr 2024

深度神经网络中的自组织与 $1/f$ 噪声

Nicholas Chong Jia Le*Department of Physics
National University of Singapore, 117551 Singapore

Feng Ling†Complex Systems Group, Systems Science Department
Institute of High Performance Computing, A*STAR,
138632 Singapore andDepartment of Physics
National University of Singapore, 117551 Singapore（日期：2024年4月2日）

$1/f$ 噪声也称为粉红噪声，它在生物神经网络中是一种公认的现象，被认为在大脑的信息处理中发挥着重要作用。本研究发现，在以自然语言训练的深度神经网络中也存在这种 $1/f$ 噪声，其表现与生物对应物相似。具体而言，我们在“IMDb”人工智能基准数据集上训练了长短期记忆网络（LSTM），然后测量神经元激活。对不同神经元时间序列进行的去趋势波动分析（DFA）显示出明显的 $1/f$ 模式，而LSTM输入的时间序列中则不存在这种模式。有趣的是，当神经网络容量过剩、拥有足以完成学习任务的多余神经元时，激活模式会偏离 $1/f$ 噪声，转而趋向白噪声。这是因为许多神经元并未得到有效利用，在接收输入数据时几乎没有涨落。我们进一步考察了LSTM单元中“内部”和“外部”激活的 $1/f$ 噪声指数值，发现其指数变化与人脑功能性磁共振成像信号中的指数变化有些相似。我们的研究结果进一步支持了这样的假设： $1/f$ 噪声是最优学习的标志。随着深度学习模型在某些任务上接近或超越人类表现，且比其生物对应物更易于“实验研究”，本研究表明，这些模型是理解 $1/f$ 噪声基本起源的理想对象。

I. 引言

粉红噪声，也称为 $1/f$ 噪声，在许多生命系统中广泛存在 [1]。它指的是时间信号中功率谱密度与频率之间的幂律关系，其能量随频率增加而反比递减（$f^{-\beta}$），且这一关系在很宽的频率范围内成立 [2]。在生物系统中，这类噪声已在多种信号中被观测到，包括心率涨落 [3, 4], 、人类步态 [5, 6], 、人眼细胞 [7], 以及人脑活动扫描 [8]。

具体而言，在人脑中，脑电图（EEG）和功能性磁共振成像（fMRI）等活动信号 [9–11] 已发现此类模式。与健康人脑相比，患有神经系统疾病或处于衰老状态的大脑，其噪声 $1/f$ 表现有所不同 [12–15]。由于过去这种信号 $1/f$ 被视为无关噪声，因此对这种形式的脑活动的研究尚处于早期阶段；不过，有研究提出，大脑中的噪声 $1/f$对调节功能十分重要，并具有其他认知用途 [16]。许多研究测量了脑电图的标度指数 β，并得出了相似的结果 [17, 18]。 [17]中汇总的结果得出标度指数 $\beta = 1.33 \pm 0.19$，表明 $1/f$ 噪声。在功能性磁共振成像研究中 [19, 20], ，测得的标度指数小于脑电图研究中的数值，平均标度指数为 $\beta = 0.84$。当大脑执行任务时，这一指数会进一步降低，全脑平均为 $\beta = 0.72$ 。

解释 $1/f$ 噪声的最初尝试，是提出自组织临界性（SOC）的概念 [2], ：简单的沙堆模型能够通过沙粒级联雪崩产生此类涨落。后来研究发现，大脑中的神经雪崩呈现类似的幂律分布，其临界分支参数接近 1 [21], ，且幂律指数值与 SOC 模型的平均场结果一致 [22]。随后，更多研究关注了大脑的同步现象 [23]。结合模拟研究 [24–27], ，研究表明， $1/f$噪声与临界同步态有关。

尽管人们已对生物大脑中的 $1/f$ 噪声进行了广泛研究，但尚未明确证实人工神经网络（NN）中是否存在这种现象。人工神经网络是一类受其生物对应物启发、但仅在宽泛意义上受到启发的机器学习模型 [28]。人工神经网络在许多领域都取得了令人瞩目的表现，其中最强大的模型由非常深的层次结构构成，拥有大量参数和高度复杂的结构 [29, 30]。然而，从理论角度来看，其成功背后精确的学习动态和计算原理仍鲜为人知。另一方面，与生物大脑相比，这些“人工大脑”可以成为更理想的实验对象

* chong.jiale.nicholas@u.nus.edu
† fengl@ihpc.a-star.edu.sg

，因为我们可以在尽可能低的层级对它们进行控制和观察，而不受外部混杂因素的影响。因此，它们是理想的研究对象，有助于深入理解$1/f$ 噪声是否存在于其中。在这项工作中，我们考察了常见的深度学习模型中是否存在 $1/f$ 噪声。该模型用于时间序列数据建模，称为长短期记忆（LSTM）网络 [31]，并在真实世界数据集上进行训练。

II. 方法

针对不同任务，有许多成功的深度学习模型，例如专用于图像识别的 ResNet [29]，以及用于自然语言处理的基于注意力的模型 [30]。然而，要研究 $1/f$ 噪声是否存在，我们需要一种专门用于对任意长度时间序列数据建模的网络，例如循环神经网络（RNN）。RNN 是一种网络，它通过将节点的输出反馈给这些相同的节点，在时间序列中保留输入的状态。这与前馈网络不同，前馈网络中的数据只在层与层之间流动。通过这种循环机制保留先前输入的信息，RNN 在处理序列数据方面比简单的前馈网络更为擅长。

图1展示了最基本形式的 RNN，说明这类循环网络在时间维度上具有深度，这与简单前馈网络在层 维度上的深度形成对比。然而，这也意味着简单 RNN 面临与深度较大的深度神经网络相同的问题——梯度消失问题。对于超过数十个时间步的特别长的输入序列，RNN 很难收敛。从抽象层面来说，RNN 与人脑有些相似，因为人脑会持续接收信息，并通过我们的生物神经网络对其进行处理。

RNN有许多变体，范围从完全循环网络、门控循环单元网络[32], 双向循环神经网络 [33], 到物理学界更常用的储备池计算机 [34]。

在本研究中，我们使用一种称为长短期记忆（LSTM）网络的特定类型的RNN，这得益于其出色的实际应用性能 [35]。在这种架构中，创建了一个LSTM单元，用于替代图1所示普通RNN中的循环单元，以通过维持内部细胞状态 $\boldsymbol{c}$来解决梯度消失问题 [31] 。

一个长短期记忆网络单元（图2），与普通RNN单元中的单个激活相比，它包含许多不同的激活。这个长短期记忆网络单元与普通RNN单元一样，将 $\boldsymbol{x}_t$ 和 $\boldsymbol{h}_{t-1}$ 作为输入，同时还额外接收上一个细胞状态 $\boldsymbol{c}_{t-1}$作为输入。与普通RNN一样，长短期记忆网络单元也会输出当前隐藏状态 $\boldsymbol{h}_t$ ，并将当前细胞状态 $\boldsymbol{c}_t$ 传递给自身，以供下一个时间步使用。内部

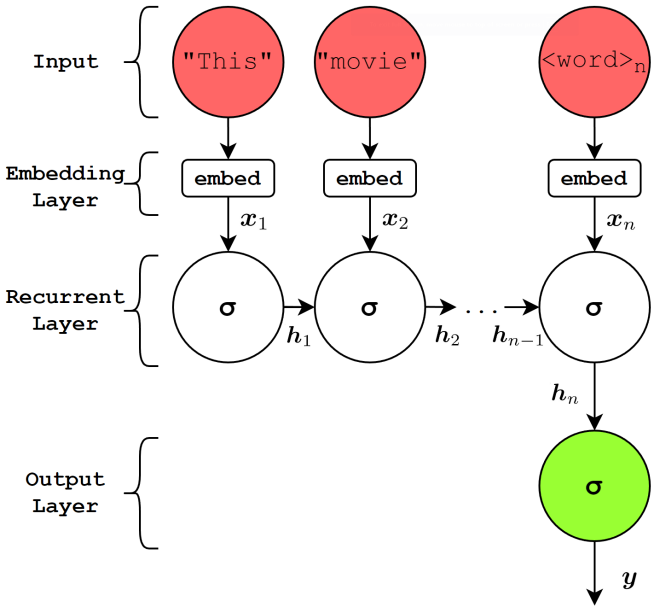


图1：以时间展开表示形式呈现的一个（多对一）循环神经网络。一个时间序列（在本例中为包含 n 个词的电影评论）按顺序输入网络。对于每个时间步t，第 t个词进入嵌入层，该层

将 t 转换 h^e wor d into a vector us ing a l_earne d 连续向量空间的表示。向量$\boldsymbol{x}_t$ 随后传入循环层，该层同时接收 $\boldsymbol{x}_t$ 和自身在前一个时间步的输出 $\boldsymbol{h}_{t-1}$。接着，循环层将其输出$\boldsymbol{h}_t$传入自身，以供下一个时间步使用。在最终时间步 n，循环层将其输出 $\boldsymbol{h}_n$ 传入

输出层，将其转换为输出 $\boldsymbol{y}$。

细胞状态 $\boldsymbol{c}$ 有助于保留时间相关性 [31]。

以下方程描述了长短期记忆网络在时间步 t进行前向传播时的不同内部状态向量：

$$
\begin{aligned}
\boldsymbol{i}_t &= \sigma_s(\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i) && (1)\\
\boldsymbol{f}_t &= \sigma_s(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f) && (2)\\
\boldsymbol{o}_t &= \sigma_s(\boldsymbol{W}_{xo}\boldsymbol{x}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o) && (3)\\
\widetilde{\boldsymbol{c}}_t &= \sigma_t(\boldsymbol{W}_{xc}\boldsymbol{x}_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c) && (4)\\
\boldsymbol{c}_t &= \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \widetilde{\boldsymbol{c}}_t && (5)\\
\boldsymbol{c}_{\text{outt}} &= \sigma_t(\boldsymbol{c}_t) && (6)\\
\boldsymbol{h}_t &= \boldsymbol{o}_t \odot \boldsymbol{c}_{\text{outt}} && (7)
\end{aligned}
$$

其中，$\odot$ 是逐元素乘积，σ_s 是 Sigmoid 激活函数 $\frac{1}{1+e^{-x}}$，σ_t 是双曲正切激活函数，$\boldsymbol{W}$ 包含模型的权重矩阵。

表 I：为 LSTM网络 选定的超参数

嵌入层大小	训练批次大小	丢弃率	L2 正则化因子	学习率
8	128	0.1	0.001	0.005

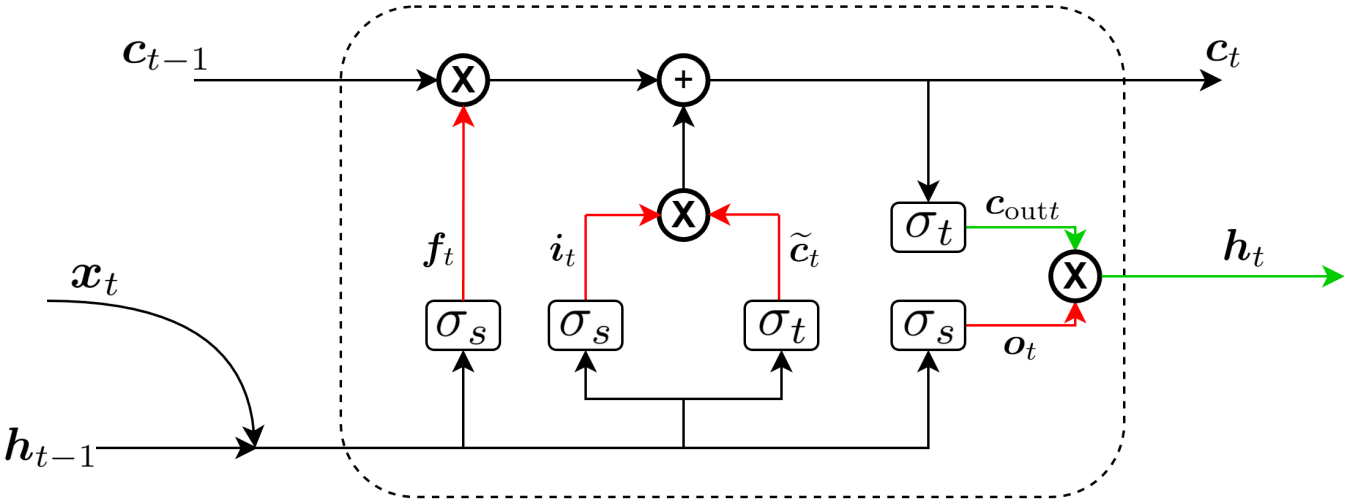


图 2：展示内部结构的 LSTM单元（虚线圆圈）。红线表示“内部”激活，绿线表示“外部”激活。 σ_t 和 σ_s 分别表示双曲正切激活和 Sigmoid 激活。

A. 数据集

情感分析任务所选用的数据集为大型电影评论数据集 [36]，该数据集包含从互联网电影数据库（IMDb）获取的 50000 条带有极性标签的电影评论 [37]。该数据集包含 25000 条正面评论（得分为 ≥ 7，满分为 10 分）和 25000 条负面评论（得分为 ≤ 4，满分为 10 分）。

我们进行了预处理，例如删除标点符号并将单词转换为小写。我们还将单词转换为词元：将排名前 4000 的单词（占全部词汇表的 88.3%）转换为唯一词元，其余单词则转换为单个 [UNK]词元。

B. 长短期记忆网络架构

长短期记忆网络由三层组成：嵌入层（将单词转换为低维内部表示向量）、LSTM 层，以及由单个神经元构成的输出层。该神经元采用 Sigmoid 激活，输出一个值，用于表示评论是正面（$y \geq 0.5$）还是负面（$y < 0.5$）。

IMDb 数据集通过 Keras [38]datasets 应用程序编程接口 (API) 获取，预处理则使用自定义代码 [39]完成。网络使用 Keras 进行训练，采用 TensorFlow 2.6.0 [40] 后端运行于 GeForce GTX 1080 图形处理器上，预处理步骤则由 Ryzen 9 3900X 中央处理器执行。

LSTM网络所使用的超参数如表 I 所示。这些超参数使用 KerasTuner [41] 库，通过 Hyperband[42] 搜索算法进行选择，共进行了 10 次 Hyperband 迭代。总体而言，在本研究中，我们遵循了针对长短期记忆网络模型的最先进实践。

C. 测量 $1/f$ 噪声

为了测量 LSTM 单元中功率谱的指数，必须获取特定激活的时间序列。为获取 Keras LSTM 单元的内部激活，我们使用带有 NumPy [43]的原生 Python 重新创建了该单元。相关代码可在 [44]处获取。随后使用去趋势波动分析（DFA） [45] 对时间序列进行分析，以获得该序列的标度指数。在本研究中，我们使用一种改进版算法，即无偏去趋势波动分析 [46]，以考虑由长度为 500 的评论所导致的相对较短的时间序列。

然后，获取任意特定激活函数指数的步骤如下（本例中为 $\boldsymbol{f}$）：

1. 将评论输入 LSTM 层，并记录每个时间步 t 上每个 LSTM 单元对应的遗忘门向量 $\boldsymbol{f}_t$，从而形成

8 个分别对应 8 个 LSTM 单元的激活时间序列。

2. 将该层各单元的激活值相加，得到 LSTM 层的总激活值 $S(t)$。3. 使用去趋势波动分析（DFA）计算激活值的涨落 $F(t)$。4. 计算 $F(t)$ 相对于 t 的双对数图斜率，以获得 DFA 标度指数 α。5. 计算功率谱指数 $\beta = 2\alpha - 1$。

III. 结果与讨论

在选择网络的最佳训练轮次时，选取网络损失最低的训练轮次。各网络取得的准确率都很高 [47]，在测试数据上的预测准确率介于 87.61% 至 89.63% 之间。

A. 测试集的指数 β

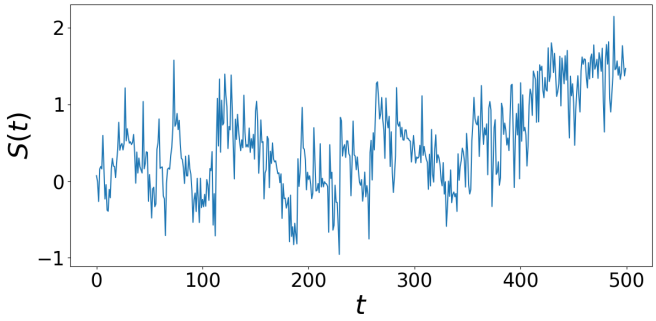
对长度为 ≥ 500 的测试数据集中的评论执行了第 II C 节所述步骤，以消除填充的影响，因为重复的相同填充词元会导致指数降低。请注意，为符合人工智能领域公认的实践，LSTM 的训练是在不考虑评论词数的情况下进行的。图3(a)显示了其中一个网络对一条评论的激活值。图3(b)是涨落 $F(t)$ 随时间变化的双对数图，其中绘有一条梯度为 1 的参考线，表示 $\beta = 2\alpha - 1 = 1$。图3(c)展示了同一网络对所有长度为 ≥ 500 的测试评论所得到的 $\boldsymbol{h}$ 指数直方图，其均值为 $\mu_\beta = 0.833 \pm 0.132$。

B. 排除输入中的 $1/f$ 噪声

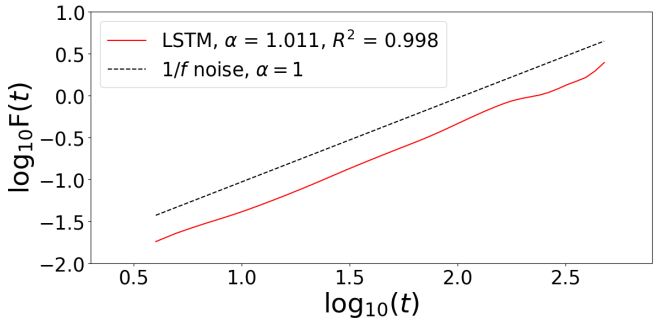
出现 $1/f$ 噪声的一种可能性是，输入 $\boldsymbol{x}$ 具有 $1/f$ 信号。因此，如果要证明LSTM中产生了 $1/f$ 噪声，就必须排除这一影响。为了确定输入数据的指数，采用了 II C中的相同过程，但使用嵌入向量代替激活向量。

LSTM的输入（嵌入后）的指数分布也绘制于图3中，显示出实际上不相关的白噪声，其均值为 $\mu_\beta = 0.002 \pm 0.116$。

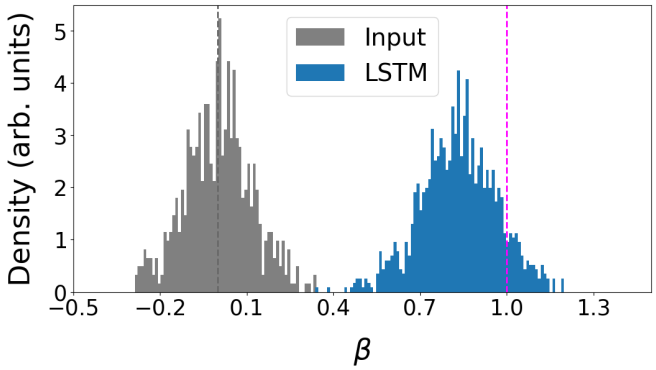
图4是对图3所示直方图的散点图，展示了激活指数与输入指数之间缺乏相关性



(a)



(b)



(c)

图3：（a）一篇901词评论（截断至500词）中，整个LSTM 层在单个样本上的激活 $\boldsymbol{h}_t$ 值。（b）对（a）中所示激活进行去趋势波动分析得到的涨落 $F(t)$的双对数图（实线，红色），并绘制了 $1/f$ 噪声的参考线（虚线，黑色）。（c）对所有长度为 ≥ 500的测试评论进行去趋势波动分析得到的指数 α 所对应的指数 β 直方图。输入嵌入激活（灰色）和 $\boldsymbol{h}_t$ （蓝色）的结果均绘出。绘制对应于 $\beta = 0$ 和 $\beta = 1$ 的点线作为参考。

数值为 0.061。这进一步支持了我们的假设：在 LSTM网络中观察到的 $1/f$ 噪声是网络固有的，而非网络输入中的 $1/f$ 噪声所导致的。

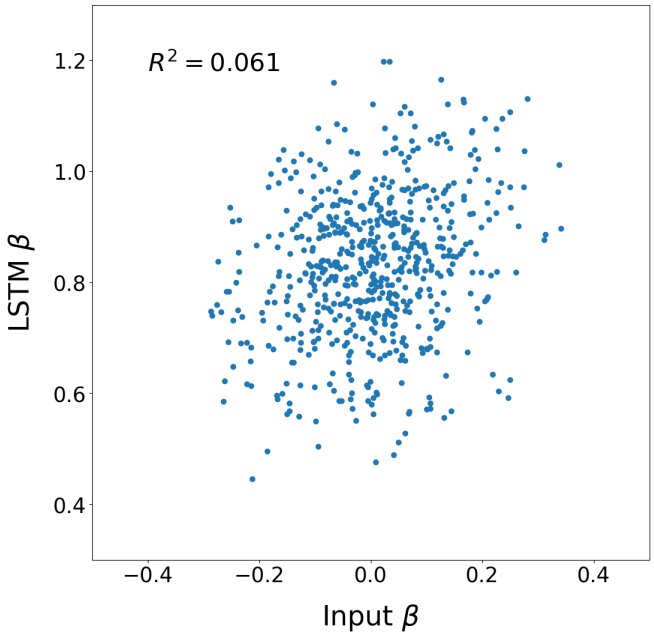


图4：测试评论中， $\boldsymbol{h}_t$ 的指数与输入 $\boldsymbol{x}_t$ 的指数的散点图，评论长度为≥ 500。使用的模型与图3相同。这里的输入指数与激活指数并不相关，这表明激活值中的 $1/f$ 现象并非源自输入中的类似模式。

C. $1/f$ 噪声的条件

受大脑这一类比以及功能正常的大脑中存在 $1/f$ 噪声的启发，我们研究了改变 LSTM 层容量时该层的指数变化。LSTM 层中的单元数量越多，其存储有用信息的容量就越大。图 5 展示了不同单元规模 n 下各 LSTM网络的 β 汇总值；这些数值汇总自多个单元规模相同、使用不同随机初始化并在同一数据集上训练的 LSTM网络。我们观察到，随着网络容量增加，指数 β 呈下降趋势。当单元数量达到 128 时，LSTM 层的汇总激活实际上趋近于白噪声。尽管 β呈下降趋势，不同网络的平均准确率基本保持不变，如表II所示。附录 A 收录了 LSTM 层内其他激活的数据，这些激活也呈现出类似趋势。

分析 LSTM 层中的各个单元时，可以发现这可能是一个原因。图6展示了最小的 LSTM 层（8 个单元）和最大的 LSTM 层（ 128 个单元）中前 8 个单元处理同一条评论时的激活 $\boldsymbol{h}_t$ 。我们清楚地看到，大型层中的多个单元似乎处于“死亡”状态，几乎没有任何活动。附录 B 展示了

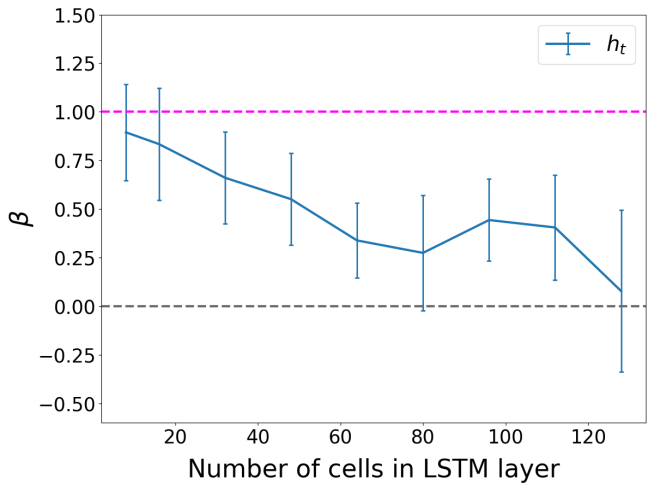


图5：激活 $\boldsymbol{h}_t$ 的 β 聚合值与 LSTM 层中单元数量的关系。图中给出了 $1/f$ 噪声（粉色虚线）和白噪声（黑色虚线）作为参考。

其他规模的 LSTM 层在相同评论下的对应图表，展示了随着容量增加，层的活动逐渐减弱。

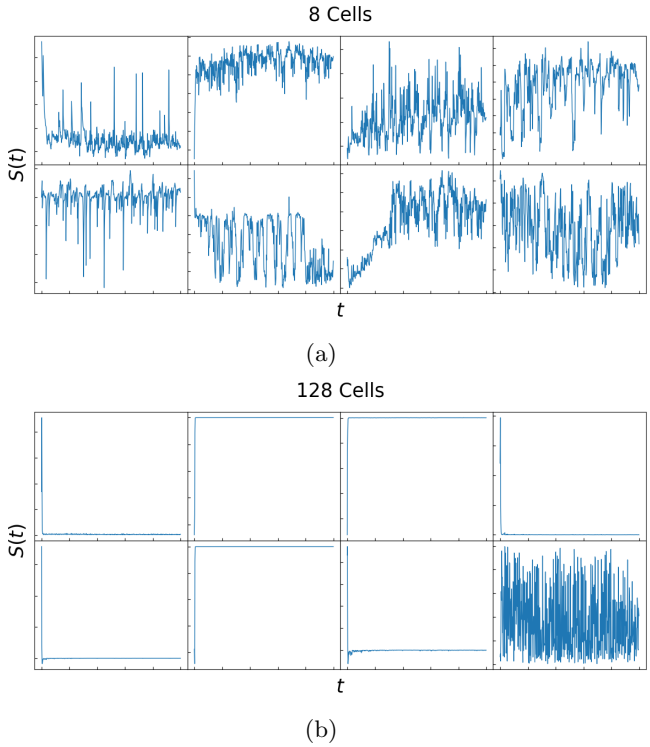


图 6： $\boldsymbol{h}_t$ 在包含 8 个单元（a）和 128 个单元（b）的 LSTM 层中对同一条评论的单元级激活。这里展示了 128 个单元的层中的前 8 个单元。我们观察到，在较大的 LSTM 层中，多个单元缺乏活动。

表II：不同规模 LSTM 网络的平均准确率

n	8	16	32	48	64	80	96	112	128
平均准确率（%）	88.95	88.97	88.78	88.79	88.96	88.56	88.32	88.78	88.24

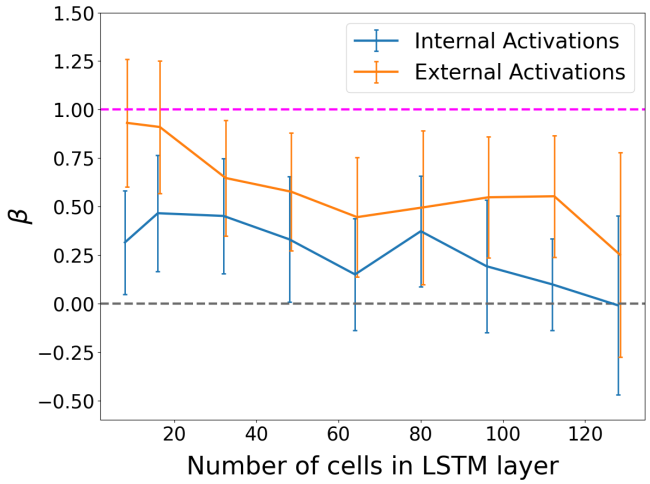


图7：外部激活（$\boldsymbol{h}_t$, $\boldsymbol{c}_{\text{out}}$）和内部激活（$\boldsymbol{i}_t$, $\boldsymbol{f}_t$, $\tilde{\boldsymbol{c}}_t$, $\boldsymbol{o}_t$）的 β 聚合值与 LSTM 层中单元数量的关系。外部激活对应的曲线略微水平移位，以便更清楚地显示误差棒。图中还给出了 $1/f$ 噪声（粉色虚线）和白噪声（黑色虚线）作为参考。

D.“内部”和“外部”激活

在对人脑中 $1/f$ 噪声的测量中，脑电图测量测量的是大脑表面的神经元，而功能性磁共振成像测量测量的是大脑体积内的神经元。这两种测量所得的激活指数明显不同，整体指数低于表面指数，不过两种测量中 β 的数值都接近 1。在我们的实验中，我们也观察到指数值有类似的分化：我们称之为“内部”激活的部分用于调节层输出，而“外部”激活则包含层输出的数值。

图 7 展示了不同规模的同一组 β LSTM网络中内部和外部激活的汇总值。我们在这两种情况下都观察到与先前相同的趋势：网络容量增加时，β的值会降低。不过，我们还看到两类激活的 β 明显不同：内部激活的 β 始终小于外部激活。如此类似于脑信号的模式，或许并非巧合，值得进一步研究。

IV.结论

在这项研究中，我们发现了令人信服的证据，表明 $1/f$ 深度神经网络（例如长短期记忆（LSTM）网络）中存在噪声。这种模式已被证明不受呈现白噪声特征的输入数据影响。由于输入数据由我们大脑处理的真实世界自然语言句子构成，我们的研究结果表明，性能接近人类水平的人工神经网络表现出与其生物学对应物非常相似的 $1/f$ 模式[10, 19, 20]。有趣的是，这种类比也适用于 LSTM 内部“内部”和“外部”神经元的噪声指数趋势，分别与功能性磁共振成像和脑电图指数相比较 [17, 20]。内部激活的幂律谱指数通常较小 β $f^{-\beta}$。

然而，值得注意的是，当神经网络的容量超过特定学习任务所需的容量时，$1/f$ 噪声模式便会消失。具体而言，当我们增加神经元数量，却未观察到模型准确率明显提升时，会发现许多神经元对输入波动没有响应，实际上并未参与该任务的“思考”，而是表现出很低甚至没有活动。这些未激活神经元的大量存在，实际上降低了神经元涨落聚合值的指数 β 。

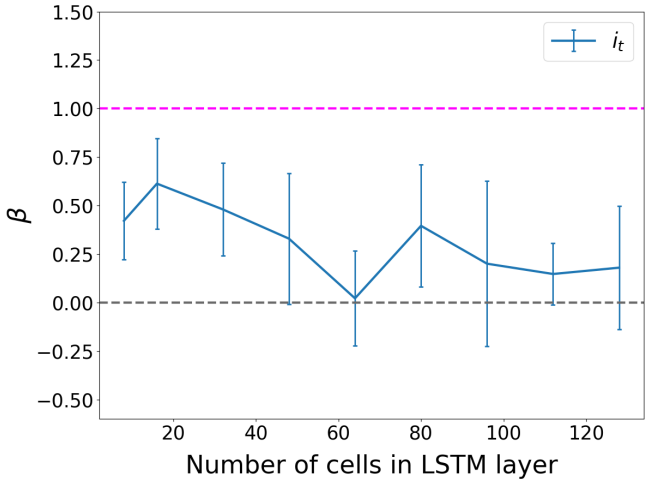
令人好奇的是，尽管生物神经网络与人工神经网络在微观细节上存在巨大差异，但它们呈现出的宏观 $1/f$ 噪声模式却惊人地相似。这种相似性表明，支配它们健康运行的背后可能存在某些更深层的原理；这些原理似乎独立于具体的神经相互作用，因为大脑神经网络与长短期记忆网络截然不同。人工神经网络不像其生物对应物那样难以通过实验进行操控和观察，因此更具“透明性”，是未来理解 $1/f$ 噪声起源的理想替代对象，也可能成为我们从 $1/f$ 噪声的视角进一步认识大脑健康运行机制的工具。

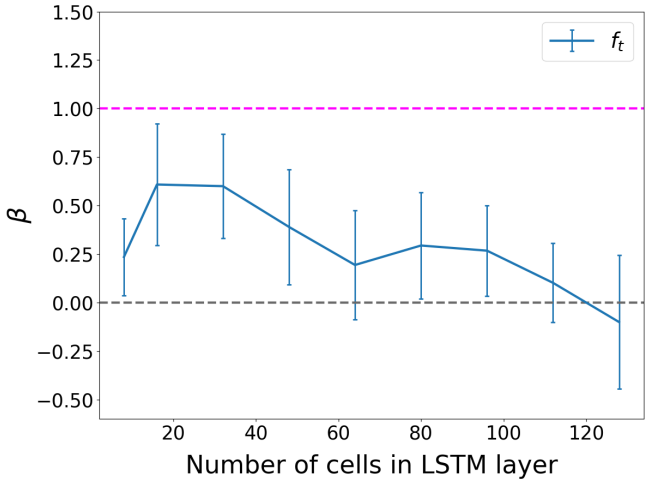
总的来说，我们的研究结果为探索以 1/f噪声为基础的生物和人工神经网络的运行状态及其背后的计算原理，开辟了新的跨学科途径。

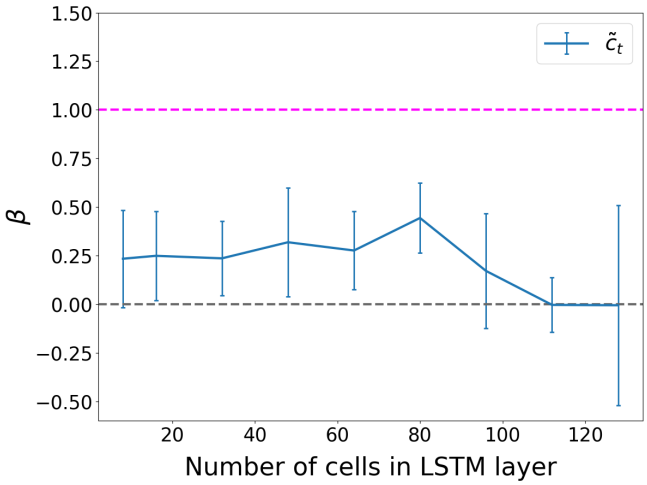
V. 附录

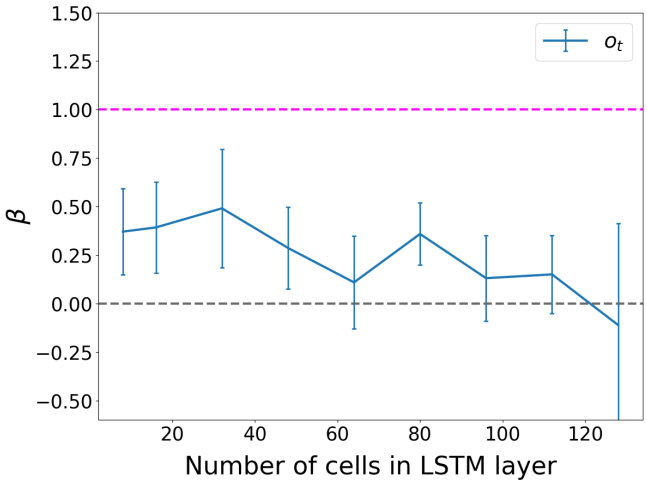
附录 A：指数与单元格数量（其他激活）

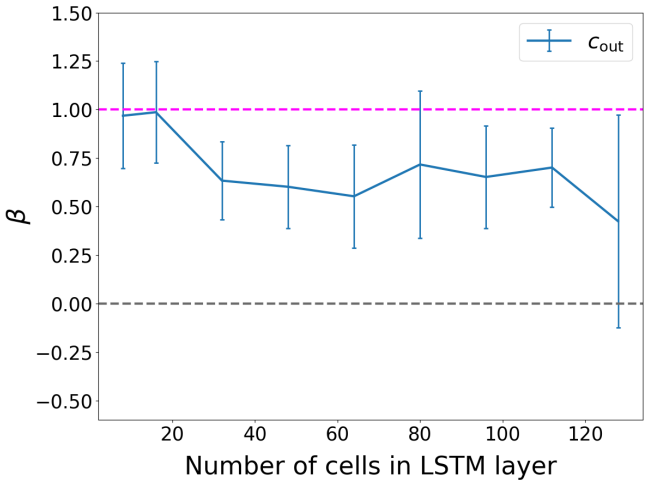
以下各图展示了 LSTM 层中其他激活所呈现的趋势，与图 5 中的趋势相同。误差棒表示 1 个标准差； $1/f$噪声（粉色虚线）和白噪声（黑色虚线）作为参考一并显示。







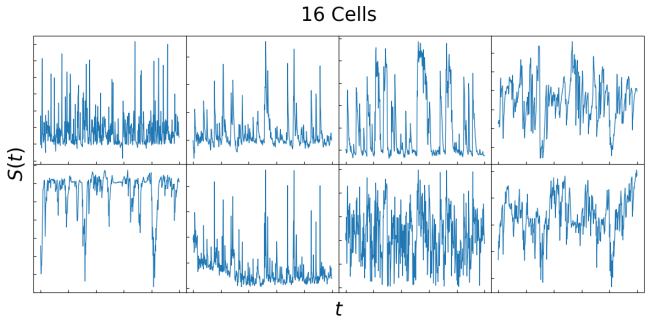


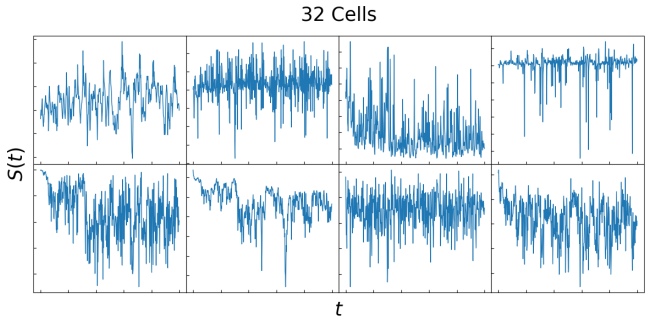


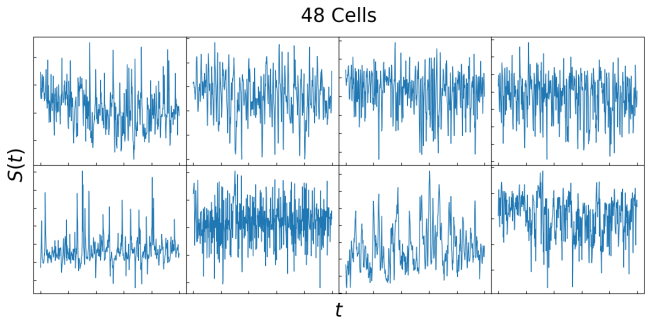
附录 B：单元级激活

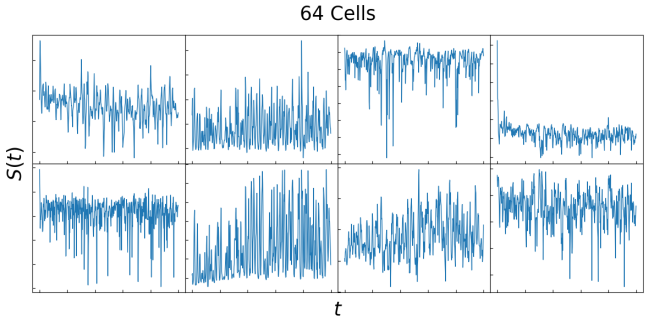
以下图示展示了不同规模 LSTM 层的单元级激活，与图

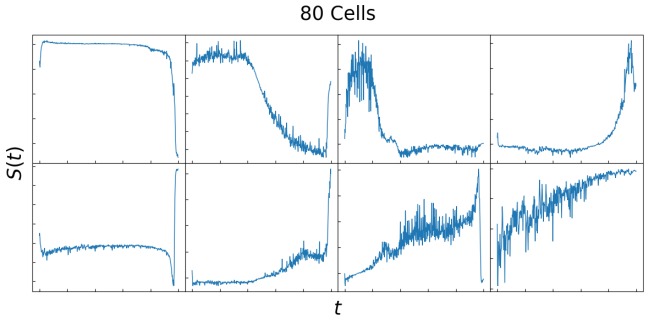
6 类似。每幅图均展示 LSTM 层中 8 个单元的激活 $\boldsymbol{h}_t$ 。所有图中的 LSTM 层接收相同的评论。我们观察到，随着 LSTM 层规模增大，部分单元的活动减弱，这与前文所述的现象一致。

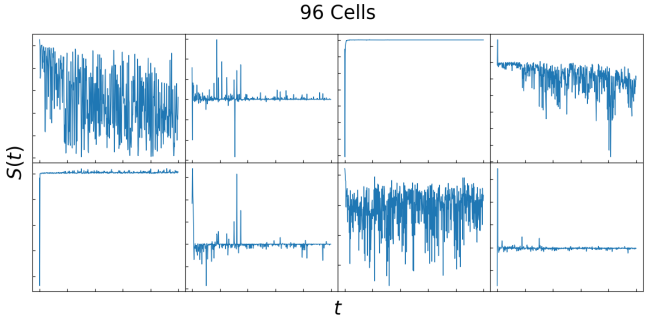


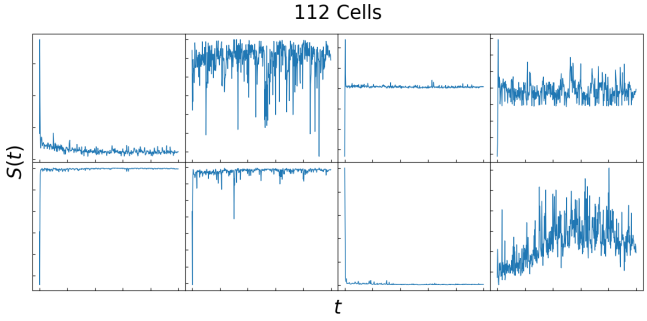












[1] M. A. Mu˜noz, 《现代物理评论》**90** (2018), 10.1103/RevModPhys.90.031001. [2] P. Bak, 《物理评论快报》**59**, 381 (1987). [3] C.-K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, and A. L. Goldberger, 《物理评论快报》**70**, 1343 (1993).

[4] Z. R. Struzik、J. Hayano、S. Sakata、S. Kwak 和 Y. Yamamoto, 《物理评论 E》**70**，050901（2004）。[5] J. M. Hausdorff、C. K. Peng、Z. Ladin、J. Y. Wei 和 A. L. Goldberger, 《应用生理学杂志》**78**，349 358 1995ˆa ( )。[6] J. M. Hausdorff、S. L. Mitchell、R. Firtion、C. K. Peng、M. E. Cudkowicz、J. Y. Wei 和 A. L. Goldberger,

《应用生理学杂志》82，262ˆa269（1997）。[7] J. M. Medina 和 J. A. D´ıaz，美国光学学会杂志 A 29，A82（2012）。[8] J. O'Byrne 和 K. Jerbi，神经科学趋势 45，820（2022）。[9] K. Linkenkaer-Hansen、V. Nikouline、J. Palva 和 R. Ilmoniemi，神经科学杂志；神经科学杂志 21，1370（2001）。[10] E. Novikov、A. Novikov、D. Shannahoff-Khalsa、B. Schwartz 和 J. Wright，《物理评论 E》56，R2387（1997）。[11] E. Niedermeyer，*Electroencephalography: Basic Principles Clinical Applications and Related Fields* 7 ，，，第 th 版，由 D. L. Schomer 和 F. H. Lopes da Silva 编（牛津大学出版社，2017）。[12] K. Linkenkaer-Hansen、S. Monto、H. Ryts¨al¨a、K. Suomi- nen、E. Isomets¨a 和 S. K¨ahk¨onen，神经科学杂志；神经科学杂志 25，10131（2005）。[13] T. Montez、S.-S. Poil、B. F. Jones、I. Manshanden、J. P. A. Verbunt、B. W. van Dijk、A. B. Brussaard、A. van Ooyen、C. J. Stam、P. Scheltens 和 K. Linkenkaer-Hansen，美国国家科学院院刊 106，1614ˆa1619（2009）。[14] F. S. Racz、K. Farkas、O. Stylianou、Z. Kaposzta、A. Czoch、P. Mukli、G. Csukly 和 A. Eke，脑与行为 11，e02047（2021）。[15] B. Voytek、M. Kramer、J. Case、K. Lepage、Z. Tempesta、R. Knight 和 A. Gazzaley，神经科学杂志；神经科学杂志 35，13257（2015）。[16] B. J. He、J. M. Zempel、A. Z. Snyder 和 M. E. Raichle，神经元 66，353（2010）。[17] N. Dehghani、C. B´edard、S. Cash、E. Halgren 和 A. Destexhe，计算神经科学杂志 29，405 2010( )。[18] A. Pathania、M. Schreiber、M. W. Miller、M. J. Euler 和 K. R. Lohse，国际心理生理学杂志 160，18（2021）。[19] B. J. He，认知科学趋势 18，480（2014）。[20] B. J. He，神经科学杂志 31，13786（2011）。[21] J. Beggs 和 D. Plenz，神经科学杂志；神经科学杂志 23，11167（2003）。[22] K. Christensen，《物理评论 E》48，3361（1993）。[23] P. Expert、R. Lambiotte、D. R. Chialvo、K. Christensen、H. J. Jensen、D. J. Sharp 和 F. Turkheimer，英国皇家学会界面杂志 8，472ˆa479（2010）。[24] S.-S. Poil、R. Hardstone、H. Mansvelder 和 K. Linkenkaer-Hansen，神经科学杂志；神经科学杂志 32，9817（2012）。[25] M. Girardi-Schappo、G. S. Bortolotto、J. J. Gonsalves、L. T. Pinto 和 M. H. R. Tragtenberg，科学报告 6（2016），10.1038/srep29561。[26] F. Lombardi、H. J. Herrmann 和 L. de Arcangelis，混沌：非线性科学跨学科杂志 27（2017），10.1063/1.4979043。[27] D. Aguilar-Vel ˜A!'zquez 和 L. Guzm ˜A!'n-Vargas，科学报告 9（2019），10.1038/s41598-018-37920-w。[28] C. C. Aggarwal，*Neural Networks and Deep Learning*（施普林格，2018）。[29] K. He、X. Zhang、S. Ren 和 J. Sun，载于 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016 （）。[30] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、L. u. Kaiser 和 I. Polosukhin，载于 *Advances in Neural Information Processing Systems*，第 30 卷，由 I. Guyon、U. V. Luxburg、S. Bengio、H. Wallach、R. Fergus、S. Vishwanathan 和 R. Garnett 编（Curran Associates, Inc., 2017）。[31] S. Hochreiter 和 J. Schmidhuber，《神经计算》**9** (1997), 10.1162/neco.1997.9.8.1735。[32] K. Cho、B. van Merrienboer、C. Gulcehre、D. Bahdanau、F. Bougares、H. Schwenk 和 Y. Bengio，(2014), 1406.1078。[33] M. Schuster 和 K. Paliwal，IEEE 信号处理汇刊**45**, 2673ˆa2681 (1997)。[34] G. Tanaka、T. Yamane、J. B. H´eroux、R. Nakane、N. Kanazawa、S. Takeda、H. Numata、D. Nakano 和 A. Hirose，《神经网络》**115**, 100 (2019)。[35] I. Sutskever、O. Vinyals 和 Q. V. Le，载于 *Advances in Neural Information Processing Systems* 27， ，由 Z. Ghahramani、M. Welling、C. Cortes、N. Lawrence 和 K. Weinberger 编（Curran Associates, Inc., 2014）。[36] A. L. Maas、R. E. Daly、P. T. Pham、D. Huang、A. Y. Ng 和 C. Potts，载于 *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*（Association for Computational Linguistics，波特兰，俄勒冈州，美国，2011）第 142–150 页。[37] Https://www.imdb.com/。[38] F. Chollet *et al.*，"Keras"，https://keras.io (2015)。[39] https://github.com/NicholasCJL/imdb-LSTM/blob/ master/data_processing.py。[40] M. Abadi、A. Agarwal、P. Barham、E. Brevdo、Z. Chen、C. Citro、G. S. Corrado、A. Davis、J. Dean、M. Devin、S. Ghemawat、I. Goodfellow、A. Harp、G. Irving、M. Is- ard、Y. Jia、R. Jozefowicz、L. Kaiser、M. Kudlur、J. Levenberg、D. Man´e、R. Monga、S. Moore、D. Murray、C. Olah、M. Schuster、J. Shlens、B. Steiner、I. Sutskever、K. Talwar、P. Tucker、V. Vanhoucke、V. Vasudevan、F. Vi´egas、O. Vinyals、P. Warden、M. Wattenberg、M. Wicke、Y. Yu 和 X. Zheng，"TensorFlow：异构系统上的大规模机器学习"，(2015)，软件可从 tensorflow.org 获取。[41] T. O'Malley、E. Bursztein、J. Long、F. Chollet、H. Jin、L. Invernizzi、 *et al.*，"Keras Tuner"，https://github. com/keras-team/keras-tuner (2019)。[42] L. Li、K. Jamieson、G. DeSalvo、A. Rostamizadeh 和 A. Talwalkar，《机器学习研究期刊》**18**, 6765 2017( )。[43] C. R. Harris、K. J. Millman、S. J. van der Walt、R. Gom- mers、P. Virtanen、D. Cournapeau、E. Wieser、J. Tay- lor、S. Berg、N. J. Smith、R. Kern、M. Picus、S. Hoyer、M. H. van Kerkwijk、M. Brett、A. Haldane、J. F. del R´ıo、M. Wiebe、P. Peterson、P. G´erard-Marchant、K. Shep- pard、T. Reddy、W. Weckesser、H. Abbasi、C. Gohlke 和 T. E. Oliphant，《自然》**585**, 357 (2020)。[44] https://github.com/NicholasCJL/imdb-LSTM/blob/ master/get_LSTM_internal_vectorized.py。[45] C. K. Peng、S. V. Buldyrev、S. Havlin、M. Simons、H. E. Stanley 和 A. L. Goldberger，《物理评论 E》**49** (1994), 10.1103/PhysRevE.49.1685。[46] Q. Yuan、C. Gu、T. Weng 和 H. Yang，《物理学 A》**505**, 179 2018( )。[47]供参考，基于长短期记忆网络的最佳分类器准确率为 94.1%，而表现最佳的模型是 Transformer，准确率为 96.1%。数据来源：https://paperswithcode.com/sota/ sentiment-analysis-on-imdb。