[1-62]

From **Unsupervised Predictive Memory in a Goal-Directed Agent.pdf**

1. Bahdanau, D., K. Cho, and Y.J.a.p.a. Bengio, *Neural machine translation by jointly learning to align and translate.* 2014.

神经机器翻译是最近提出的机器翻译方法。与传统的统计机器翻译不同，神经机器翻译旨在构建单个神经网络，可以联合调整以最大化翻译性能。

2. BAHLDW, H., *Cerebral cortex as model builder.* 1955.

3. Bastos, A.M., et al., *Canonical microcircuits for predictive coding.* 2012. **76**(4): p. 695-711.

我们重新审视了已知的观点，即在等级皮层区域之间传递的信息实现了贝叶斯推理的一种形式 - 仔细关注神经元群体之间内在联系的含义。 通过导出用于这些计算的规范形式，可以将特定神经元群体与特定计算角色相关联。

4. Beattie, C., et al., *Deepmind lab.* 2016.

5. Blodgett, H.C.J.U.o.C.p.i.p., *The effect of the introduction of reward upon the maze performance of rats.* 1929.

6. Brady, T.F., et al., *Visual long-term memory has a massive storage capacity for object details.* 2008. **105**(38): p. 14325-14329.

记忆研究的一个主要教训是人类的记忆力是错误的，不精确的，并且受到干扰。这些结果对认知模型有影响，其中容量限制强加了主要的计算约束（例如，对象识别的模型），并且对存储器存储和检索的神经模型提出了挑战，其必须能够解释如此大的和详细的存储容量。

7. Clayton, N.S. and A.J.N. Dickinson, *Episodic-like memory during cache recovery by scrub jays.* 1998. **395**(6699): p. 272.

jays的恢复偏好表明记录了特定食物的缓存地点和时间，从而实现了非人类动物中类似情节记忆的行为标准。

8. Corbit, L.H. and B.W.J.J.o.N. Balleine, *The role of the hippocampus in instrumental conditioning.* 2000. **20**(11): p. 4233-4239.

大量证据表明，在仪器调节中，大鼠可以编码它们所暴露的特定作用 - 结果关联以及动作产生其相关结果的因果程度。三项实验评估了海马体在编码器乐学习的这些方面的参与。这些结果表明，在器械条件反射中，背侧海马的损伤选择性地损害大鼠代表动作与其后果之间的因果关系的能力。

9. Dayan, P.J.N.C., *Improving generalization for temporal difference learning: The successor representation.* 1993. **5**(4): p. 613-624.

10. Drucker, H., et al., *Boosting and other ensemble methods.* 1994. **6**(6): p. 1289-1301.

11. Duan, Y., et al. *One-shot imitation learning*. in *Advances in neural information processing systems*. 2017.

模仿学习通常被用于孤立地解决不同的任务。这通常需要仔细的特征工程或大量的样本。这远远不是我们想要的：理想情况下，机器人应该能够从任何特定任务的极少数演示中学习，并立即推广到同一任务的新情况，而无需任务特定的工程。在本文中，我们提出了一个实现这种能力的元学习框架，我们称之为一次性模仿学习。

12. Dumoulin, V. and F.J.a.p.a. Visin, *A guide to convolution arithmetic for deep learning.* 2016.

我们介绍了一个指南，以帮助深度学习从业者理解和操纵卷积神经网络架构。 该指南阐明了卷积，汇集和转置卷积层的各种属性（输入形状，核形状，零填充，步幅和输出形状）之间的关系，以及卷积和转置卷积层之间的关系。 针对各种情况导出关系，并且为了使它们直观而进行说明。

13. Eliasmith, C., et al., *A large-scale model of the functioning brain.* 2012. **338**(6111): p. 1202-1205.

14. Finkelstein, A., L. Las, and N.J.A.r.o.n. Ulanovsky, *3-D maps and compasses in the brain.* 2016. **39**: p. 171-196.

我们认为海马位置 - 细胞图在所有三个维度上都是度量的，并且可能由2-D和3-D片段组成，这些片段通过3-D头部方向单元被拼接成全局3-D度量表示。 最后，我们提出海马结构可能实现卡尔曼滤波器的神经模拟，卡尔曼滤波器是用于三维导航的标准工程算法。

15. Gemici, M., et al., *Generative temporal models with memory.* 2017.

我们考虑使用长程依赖性对时间数据建模的一般问题，其中基于时间上遥远的过去观察，新观察是完全或部分可预测的。一个足够强大的时间模型应该将序列的可预测元素与不可预测的元素分开，表达对那些不可预测元素的不确定性，并快速识别可能有助于预测未来的新元素。为了创建这样的模型，我们介绍了使用外部存储器系统增强的Generative Temporal Models。它们是在变分推理框架内开发的，它提供了一种实用的培训方法和方法，以深入了解模型的运作。我们在一系列稀疏的长期时间依赖性问题上表明，这些模型从序列的早期存储信息，并有效地重用这些存储的信息。这使得它们比基于众所周知的递归神经网络（如LSTM）的现有模型表现得更好。

16. Gluck, M.A. and C.E.J.H. Myers, *Hippocampal mediation of stimulus representation: A computational theory.* 1993. **3**(4): p. 491-516.

17. Graves, A., A.-r. Mohamed, and G. Hinton. *Speech recognition with deep recurrent neural networks*. in *2013 IEEE international conference on acoustics, speech and signal processing*. 2013. IEEE.

本文研究了深度递归神经网络，它将深度网络中证明非常有效的多层次表示与灵活使用赋予RNN的远程上下文相结合。

18. Graves, A., et al., *Hybrid computing using a neural network with dynamic external memory.* 2016. **538**(7626): p. 471.

人工神经网络非常擅长于感觉处理，序列学习和强化学习，但由于缺乏外部记忆，它们在表示变量和数据结构以及长时间存储数据的能力方面受到限制。在这里，我们介绍一种称为可微分神经计算机（DNC）的机器学习模型，它包括一个可以读写外部存储矩阵的神经网络，类似于传统计算机中的随机存取存储器。像传统计算机一样，它可以使用其内存来表示和操纵复杂的数据结构，但是，像神经网络一样，它可以从数据中学习。

19. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

更深入的神经网络更难训练。我们提出了一个残留的学习框架，以便于对比以前使用的网络深度更深的网络进行培训。我们参考层输入明确地将层重新表示为学习残差函数，而不是学习未引用的函数。我们提供全面的经验证据表明这些残留网络更容易优化，并且可以从显着增加的深度获得准确性。

20. Hermann, K.M., et al., *Grounded language learning in a simulated 3d world.* 2017.

我们越来越多地被人工智能技术所包围，这些技术代表我们做出决策并执行行动。这迫切需要与人工语言沟通，指导和引导人工代理的一般手段，这种沟通最具说服力的手段。为了以可扩展的方式实现这一目标，智能体必须能够将语言与世界和行动联系起来;也就是说，他们对语言的理解必须有基础和体现。然而，学习扎根语言是人工智能研究中一个众所周知的挑战性问题。在这里，我们介绍了一个学习在模拟3D环境中解释语言的代理，在该环境中，为了成功执行书面指令而获得奖励。通过强化和无监督学习的组合训练，并且从最小的先验知识开始，代理学习将语言符号与其物理环境的紧急感知表示和相关的动作序列相关联。代理人对语言的理解超出了以往的经验，使其能够将熟悉的语言应用于不熟悉的情境并解释完全新颖的指令。此外，随着语义知识的增长，该代理学习新单词的速度也在增加。这种用于推广和引导语义知识的工具表明了本方法用于协调模糊自然语言与物理世界的复杂性的潜力。

21. Hindy, N.C., F.Y. Ng, and N.B.J.N.n. Turk-Browne, *Linking pattern completion in the hippocampus to predictive coding in visual cortex.* 2016. **19**(5): p. 665.

预测编码框架感知的模型作为期望约束感觉表征的生成过程。 这些模型考虑了刺激如何随时间移动或变化的预期，但没有解决基于先前经验可能出现的其他不同刺激的预期。 我们表明，人类视觉皮层中基于记忆的期望与模式完成的海马机制有关。

22. Hochreiter, S. and J.J.N.c. Schmidhuber, *Long short-term memory.* 1997. **9**(8): p. 1735-1780.

23. Howard, M.W. and M.J.J.J.o.M.P. Kahana, *A distributed representation of temporal context.* 2002. **46**(3): p. 269-299.

24. Hunt, L.T. and B.Y.J.N.R.N. Hayden, *A distributed, hierarchical and recurrent framework for reward-based choice.* 2017. **18**(3): p. 172.

许多关于基于奖励的选择的说法都论证了串行和功能本地化的不同组件过程。在这篇意见文章中，我们争论另一种观点，即从多个大脑区域分布的重复计算中产生选择。 我们强调神经解剖学的几个特征可能如何支持选择的实施，包括复发神经网络中的相互抑制和跨皮质信息处理的时间尺度的层次组织。

25. Ioffe, S. and C.J.a.p.a. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift.* 2015.

训练深度神经网络很复杂，因为在训练期间每层输入的分布发生变化，因为前一层的参数发生了变化。这通过要求较低的学习速率和仔细的参数初始化来减慢训练，并且使得训练具有饱和非线性的模型变得非常困难。我们将这种现象称为内部协变量偏移，并通过归一化层输入来解决问题。我们的方法的优势在于使标准化成为模型体系结构的一部分，并为每个培训小批量执行标准化。

26. Jaderberg, M., et al., *Reinforcement learning with unsupervised auxiliary tasks.* 2016.

深层强化学习代理通过直接最大化累积奖励来实现最先进的结果。 但是，环境包含更多种类的可能训练信号。 在本文中，我们介绍了一种通过强化学习同时最大化许多其他伪奖励功能的智能体。

27. Kalman, R.E.J.J.o.b.E., *A new approach to linear filtering and prediction problems.* 1960. **82**(1): p. 35-45.

28. Kingma, D.P. and J.J.a.p.a. Ba, *Adam: A method for stochastic optimization.* 2014.

我们引入Adam，一种基于一阶梯度的随机目标函数优化算法，基于低阶矩的自适应估计。该方法实现简单，计算效率高，存储器要求少，对梯度的对角重新缩放不变，并且非常适合于数据和/或参数方面较大的问题。

29. Kingma, D.P. and M.J.a.p.a. Welling, *Auto-encoding variational bayes.* 2013.

在存在具有难治性后验分布的连续潜在变量和大型数据集的情况下，我们如何在定向概率模型中进行有效的推理和学习？ 我们引入了随机变分推理和学习算法，可以扩展到大型数据集，并且在一些温和的可微分条件下，甚至可以在难以处理的情况下工作。

30. Kok, P., J.F. Jehee, and F.P.J.N. De Lange, *Less is more: expectation sharpens representations in the primary visual cortex.* 2012. **75**(2): p. 265-270.

关于视觉世界的先前期望通过允许我们从嘈杂和模糊的数据中快速推断出合理的解释来促进感知。 这种促进的神经机制仍然很不清楚。 在这里，我们使用功能磁共振成像（fMRI）和多变量模式分析（MVPA）技术来测量人类志愿者早期视觉皮层中神经活动的幅度和代表性内容。 我们发现虽然感知期望降低了初级视觉皮层（V1）中的神经反应幅度，但它改善了该区域的刺激表现，如MVPA所揭示的。 这种信息改进与任务相关性的注意力调节无关。 最后，当预期的刺激特征相关时，V1的信息改善与受试者的行为改善相关。 这些数据表明，期望通过锐化感官表征来促进感知。

31. Lake, B.M., R. Salakhutdinov, and J.B.J.S. Tenenbaum, *Human-level concept learning through probabilistic program induction.* 2015. **350**(6266): p. 1332-1338.

学习新概念的人通常只能从一个例子中成功地进行概括，但机器学习算法通常需要数十或数百个示例才能以相似的精度执行。 人们还可以比传统算法以更丰富的方式使用学到的概念 - 用于行动，想象和解释。 我们提出了一个计算模型，它捕获了大量简单视觉概念的人类学习能力：来自世界字母表的手写字符。 该模型将概念表示为最佳解释贝叶斯标准下观察到的示例的简单程序。 在具有挑战性的一次性分类任务中，该模型实现了人类水平的表现，同时优于最近的深度学习方法。 我们还提出了几个“视觉图灵测试”，探讨模型的创造性概括能力，这在很多情况下与人类行为无法区分。

32. Luck, S.J. and E.K.J.N. Vogel, *The capacity of visual working memory for features and conjunctions.* 1997. **390**(6657): p. 279.

33. Mehri, S., et al., *Samplernn: An unconditional end-to-end neural audio generation model.* 2016.

在本文中，我们提出了一种基于一次生成一个音频样本的无条件音频生成的新模型。 我们证明了我们的模型，它从无记忆模块，即自回归多层感知器和层次结构中的状态递归神经网络中获益，能够在三个数据集上捕获非常长时间跨度的时间序列变化的潜在来源 不同性质。 对生成的样本进行人体评估表明，我们的模型优于竞争模型。 我们还展示了模型的每个组件如何有助于展示的性能。

34. Mirowski, P., et al., *Learning to navigate in complex environments.* 2016.

学习使用动态元素在复杂环境中导航是开发AI代理的重要里程碑。在这项工作中，我们将导航问题表述为强化学习问题，并表明通过利用多模态感知输入的额外辅助任务可以显着提高数据效率和任务性能。特别是我们考虑用辅助深度预测和循环闭包分类任务共同学习目标驱动的强化学习问题。 这种方法可以学习从复杂3D迷宫中的原始感官输入导航，即使在目标位置经常变化的情况下也接近人类级别的表现。 我们提供了对代理行为，其本地化能力及其网络活动动态的详细分析，表明代理隐式学习了关键导航能力。

35. Mnih, V., et al. *Asynchronous methods for deep reinforcement learning*. in *International conference on machine learning*. 2016.

我们提出了一种概念上简单的轻量级重量框架，用于深度加固学习，使用异步梯度下降来优化深度神经网络控制器。

36. Mnih, V., et al., *Human-level control through deep reinforcement learning.* 2015. **518**(7540): p. 529.

我们证明，仅接收像素和游戏分数作为输入的深度Q网络代理能够超越所有先前算法的性能，并且在49个游戏中达到与专业人类游戏测试者相当的水平，使用相同的算法，网络架构和超参数。这项工作弥合了高维度感官输入和行动之间的鸿沟，从而产生了第一个能够学习在各种具有挑战性的任务中表现出色的人工智能代理。

37. Momennejad, I., et al., *The successor representation in human reinforcement learning.* 2017. **1**(9): p. 680.

我们研究了一个中间算法族，即后继表示，它通过存储部分计算的动作值来平衡灵活性和效率：对未来事件的预测。这些预先计算策略在任务更改后更新其选择的方式不同。后继代表依赖于存储的关于未来状态的预测，预测了对任务事件序列变化不敏感的独特签名，但是随着奖励的变化灵活调整。我们在与人类进行的两项行为研究中提供了这种差异敏感性的证据。

38. Morris, R.J.J.o.n.m., *Developments of a water-maze procedure for studying spatial learning in the rat.* 1984. **11**(1): p. 47-60.

39. Moustafa, A.A., et al., *Why trace and delay conditioning are sometimes (but not always) hippocampal dependent: a computational model.* 2013. **1493**: p. 48-67.

在这里，我们提出了我们先前的海马功能和刺激表示的试验级模型的扩展，可以在统一的框架内解释这些发现。具体地，当前模型包括自适应复发附属连接，其有助于表示试验内时间信息。

40. Neisser, U., *Cognitive psychology: Classic edition*. 2014: Psychology Press.

41. O'keefe, J.J.C., *The hippocampus as a cognitive map.* 1978.

42. Oh, J., et al., *Control of memory, active perception, and action in minecraft.* 2016.

我们使用这些任务系统地比较和对比现有的深度强化学习（DRL）架构与我们新的基于内存的DRL架构。实验结果表明，我们的新架构比现有的DRL架构更好地推广到看不见的环境。

43. Pilley, J.W. and A.K.J.B.p. Reid, *Border collie comprehends object names as verbal referents.* 2011. **86**(2): p. 184-195.

总之，这些研究表明Chaser获得了对名词的指称理解，这种能力通常归因于儿童，其中包括：（a）意识到单词可以指对象，（b）意识到将语言映射到对象指示物上的语言提示，以及（c）意识到名称可以指独特的对象或对象的类别，而不依赖于针对这些对象的行为。

44. Rao, R.P. and D.H.J.N.n. Ballard, *Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects.* 1999. **2**(1): p. 79.

45. Rezende, D.J., S. Mohamed, and D.J.a.p.a. Wierstra, *Stochastic backpropagation and approximate inference in deep generative models.* 2014.

我们将来自深度神经网络和近似贝叶斯推理的思想结合起来，得出一类广义的深度定向生成模型，赋予了可扩展推理和学习的新算法。 我们的算法引入了一种识别模型来表示近似的后验分布，并作为数据的随机编码器。 我们开发了随机反向传播 - 通过随机变量进行反向传播的规则 - 并使用它来开发一种算法，允许联合优化生成和识别模型的参数。

46. Schulman, J., et al., *High-dimensional continuous control using generalized advantage estimation.* 2015.

策略梯度方法在强化学习中是一种有吸引力的方法，因为它们直接优化累积奖励，并且可以直接用于非线性函数逼近器，例如神经网络。两个主要挑战是通常需要的大量样本，以及尽管输入数据不稳定但难以获得稳定和稳定的改进。我们通过使用价值函数以牺牲某些偏差为代价来大幅减少政策梯度估计的方差，并利用类似于TD（lambda）的优势函数的指数加权估计来解决第一个挑战。

47. Seward, J.P.J.J.o.E.P., *An experimental analysis of latent learning.* 1949. **39**(2): p. 177.

48. Silver, D., et al., *Mastering the game of Go with deep neural networks and tree search.* 2016. **529**(7587): p. 484.

在这里，我们介绍了一种新的计算机Go方法，它使用“价值网络”来评估董事会的位置和“策略网络”来选择移动。这些深度神经网络通过人类专家游戏的监督学习和自我游戏的强化学习的新颖组合进行训练。在没有任何前瞻性搜索的情况下，神经网络在最先进的蒙特卡罗树搜索程序的级别上玩Go，模拟数千个随机游戏。

49. Song, H.F., G.R. Yang, and X.-J.J.E. Wang, *Reward-based training of recurrent neural networks for cognitive and value-based tasks.* 2017. **6**: p. e21492.

在这里，我们实施循环神经网络的基于奖励的训练，其中价值网络通过使用决策网络的活动来指导学习以预测未来奖励。

50. Srivastava, N., et al., *Dropout: a simple way to prevent neural networks from overfitting.* 2014. **15**(1): p. 1929-1958.

Weshow认为，Dropout可以提高神经网络在视觉，语音识别，文档分类和计算生物学方面的监督学习能力，在许多基准数据集上获得最先进的结果。

51. Sutton, R. and A. Barto, *Reinforcement Learning: An Introduction Vol. 1 MIT press*. 1998, Cambridge.

52. Sutton, R.S., et al. *Policy gradient methods for reinforcement learning with function approximation*. in *Advances in neural information processing systems*. 2000.

53. Thistlethwaite, D.J.P.b., *A critical review of latent learning and related experiments.* 1951. **48**(2): p. 97.

54. Todd, M.T., Y. Niv, and J.D. Cohen. *Learning to use working memory in partially observable environments through dopaminergic reinforcement*. in *Advances in neural information processing systems*. 2009.

我们将机器学习中关于使用有限记忆系统的想法集中在一起因此，我们提出了一个规范的门控模型，该模型通过在线时间差分方法学习使用工作记忆来最大化一般部分可观察设置中的打折未来奖励。该模型成功地解决了基准工作记忆问题，并表现出类似于那些的限制。

55. Tolman, E.C. and C.H.J.U.o.C.p.i.p. Honzik, *Introduction and removal of reward, and maze performance in rats.* 1930.

56. Tolman, E.C.J.P.r., *Cognitive maps in rats and men.* 1948. **55**(4): p. 189.

57. Tse, D., et al., *Schemas and memory consolidation.* 2007. **316**(5821): p. 76-82.

在对大鼠使用海马依赖性配对相关任务的实验中，随着时间的推移，随着推定的新皮质模式的逐渐发展，风味 - 地方关联的记忆变得持久。 仅针对一项试验训练的新痕迹随后被同化并且迅速地与海马无关。 在一次试验学习期间，模式也在创建持久的联想记忆表示中起到了因果作用。 新皮层图式的概念可以将知识结构的心理记录与系统记忆整合的神经生物学理论联系起来。

58. Wang, J.X., et al., *Learning to reinforcement learn.* 2016.

在目前的工作中，我们介绍了一种应对这一挑战的新方法，我们将其称为深度元强化学习。以前的工作表明，经常性网络可以在完全监督的环境中支持元学习。我们将此方法扩展到RL设置。出现的是使用一种RL算法训练的系统，但其重复动态实现了第二个完全独立的RL过程。第二种学习的RL算法可以以任意方式与原始算法不同。重要的是，因为它是学习的，所以它被配置为利用训练域中的结构。

59. Washburn, D.A. and J.P.J.I.J.o.C.P. Gulledge, *A species difference in visuospatial memory in adult humans and rhesus monkeys: The concentration game.* 2002. **15**(4).

60. Werbos, P.J.J.P.o.t.I., *Backpropagation through time: what it does and how to do it.* 1990. **78**(10): p. 1550-1560.

61. Weston, J., S. Chopra, and A.J.a.p.a. Bordes, *Memory networks.* 2014.

62. Wise, S.P. and E.A.J.T.i.n. Murray, *Arbitrary associations between antecedents and actions.* 2000. **23**(6): p. 271-276.