《论 AGI 的架构》

December 1, 2019 中期报告

YKY 甄景贤

Independent researcher, Hong Kong

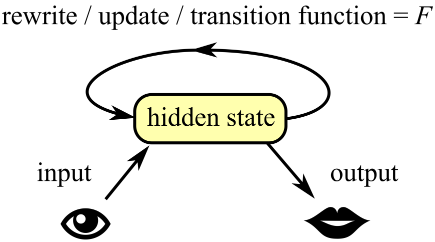
generic.intelligence@gmail.com

Table of contents

Hello 各位朋友 ☺️
近来有些进展，但只能算是「中段结果」，和大家分享一下。
也希望能找到合作者。

最简单的 AGI 架构

- 最简单的 AGI 架构是这样的：（它包含一个 recurrent **回路**）

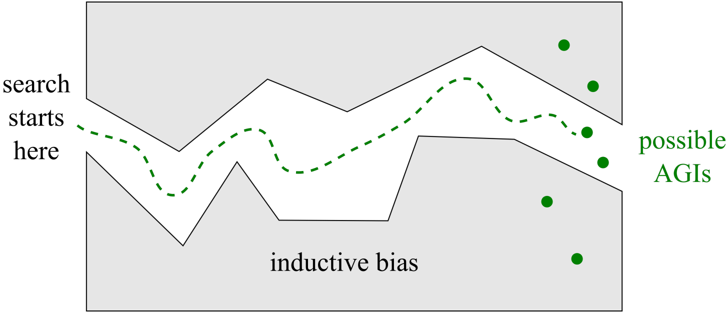


(1)

- 它在 **强化学习** 的框架下，根据 Bellman 最优化条件，将 奖励 最大化
- Transition function F 可以用一个 神经网络 实现
- 根据我先前解释过的 no free lunch 理论，这个架构的问题是缺乏 inductive bias，学习太慢

No free lunch (NFL) 的迷思

- 根据 no free lunch 哲学，没有所谓「好」与「坏」的归纳偏好
- 只要能加速学习的，但又不完全 切掉 AGI，都是好 bias

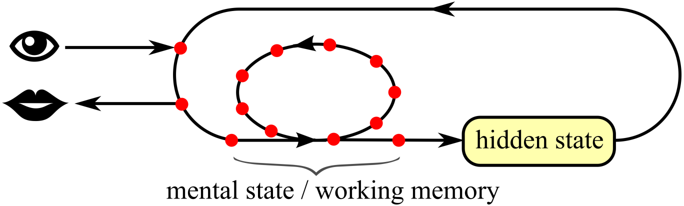


(2)

- 例如，可以将 F 的神经网络 变 **稀疏** (sparse)，但保持 深度 (deep)
- 但我之前提议用 **逻辑结构** 作为 bias，是否多此一举呢?! 😖

「双回路」架构

- 假设 working memory 是由分离的 **命题** (•) 构成的，可以提出一个「双回路」的架构：(这意思是，例如回路中载有 10 个命题，则它运转 10 次，隐状态即「浓缩」了这 10 个命题的资讯，然后输出一个新命题)

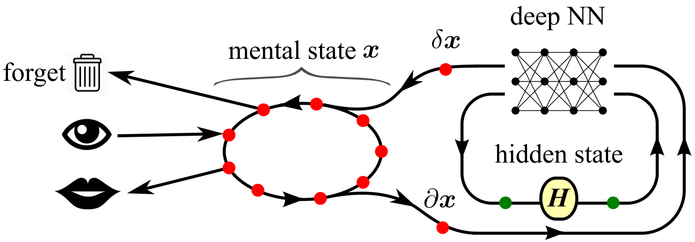


(3)

- 这种架构很可能是人脑的架构，因为比较简单，可以进化出来
- 据我所知，BERT 架构含有一个回路 / 一个隐状态（BERT 逐个输入句子的单词，然后 隐状态 再将输出的单词 逐个吐出来）
- 似乎 将 BERT 变成「双回路」，可以得到 AGI：（但我不熟悉 BERT）

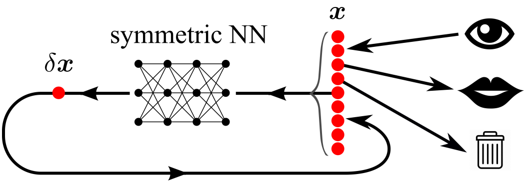
$$\boxed{\text{单词 / 概念}} \xrightarrow{\text{loop 1}} \boxed{\text{句子 / 命题}} \xrightarrow{\text{loop 2}} \boxed{\text{mental state}} \qquad (4)$$

- 图 (3) 不够严谨，详细画出来是这样的：



(5)

- 相比之下，如果用了 symmetric NN 则架构更简单：



(6)

- 问题是 (5) 和 (6) 哪个的 学习速度 更快？但不易判断 😖

强化学习 与 量子力学 之间的联系

- 强化学习 的最优化条件是 **Bellman** 方程：

$$\boxed{\text{Bellman}} \qquad S_t(x) = \max_u \{L(x,u) + \gamma S_{t+1}(x)\} \tag{7}$$

- 而 Bellman 方程 的 微分形式，是经典分析力学中的 **Hamilton-Jacobi** 方程（这点在 1970s 已经为 Kalman 和 Pontryagin 等人认识），现代叫这方程为 Hamilton-Jacobi-Bellman (HJB) 方程：

$$\boxed{\text{Hamilton-Jacobi}} \qquad \frac{\partial S(x,t)}{\partial t} = -H \tag{8}$$

- 其中 Lagrangian L 和 Hamiltonian H 之间有关系：

$$L = \text{K.E.} - \text{P.E.} \quad , \quad H = \text{K.E.} + \text{P.E.} \tag{9}$$

K.E. = kenetic energy, P.E. = potential energy.

- 至此，我们将一条 **离散**方程 变为 连续函数的**微分**方程，但没有得益

- 我最近独立发现了从经典力学的 Hamilton-Jacobi 方程 过渡到 Schrödinger 方程的 exact 形式，关键是透过 $\Psi = e^{-i\hbar S}$ 此一代入：

$$\boxed{\text{Hamilton-Jacobi}} \quad \frac{\partial S}{\partial t} = -H \quad \Rightarrow \quad i\hbar\frac{\partial \Psi}{\partial t} = H\Psi \quad \boxed{\text{Schrödinger}} \tag{10}$$

- 一直以来，量子力学的教科书 认为这种 量子化 (quantization) 过程 只能够是近似性的。后有朋友告诉我 [Field 2010] 已导出了这结果。
- 从 AI 的角度来看，这结果表示：强化学习 可以 转化为 在 Hilbert 空间中求解 Schrödinger 方程！
- 而， Schrodinger 方程可以透过 **虚数时间** (imaginary time) 转化为 热力学的 **扩散** (diffusion) 方程

$$\boxed{\text{wave eqn.}} \quad \frac{\partial \Psi}{\partial t} + i\Delta\Psi = 0 \quad \leftrightsquigarrow \quad \frac{\partial u}{\partial t} + \Delta u = 0 \quad \boxed{\text{heat eqn.}} \tag{11}$$

- 但应用到 AI 上需要 **离散化**，特别是利用 discrete Laplacian Δ 或 discrete Schrödinger operators 作用在 graph 上（状态空间是 graph）
- 暂未知这方向有没有好处 😖

Topos 理论

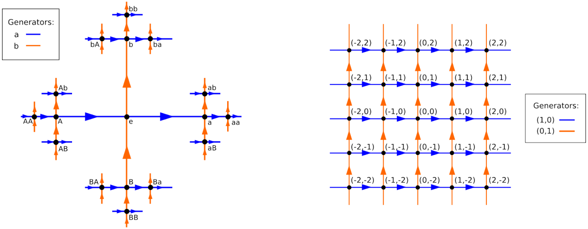
Topos 是指一个能够在里面「做逻辑」的范畴，它起源於 Lawvere 于 1950s 将 **集合论** 表述成 **范畴论** 的尝试。
在一个 topos 范畴内的物体 (objects) 可以进行三种运算：

- Cartesian product $A \times B$（对应命题逻辑的 $A \wedge B$）
- exponentiation $A \to B = B^A$ （对应命题逻辑的 $A \Rightarrow B$）
- subobject classifier $A \hookrightarrow B$ （表示 **子集** $A \subseteq B$ 的概念）

Topos 理论的重要性在於：它概括了哪些数学结构有进行 **逻辑** 的能力，例如一些 relation graphs，algebras 等。
在我的理论里，神经网络 F 负责 $U \xrightarrow{F} V$ 即 exponentiation V^U 这部分，而 U, V 是向量空间。这至少要求我们将 $A \times B$ 的结构 embed 到向量空间中。由於 $A \times B = B \times A$ 是**可交换**的，这促使我看了一下 Abel 群的理论。

交换不变性 (permutation invariance)

- 最近有個頗厉害的想法，将 Word2Vec 嵌入到 Poincaré disc / hyperbolic space [Nickel and Kiela 2017]
- 类似地，能不能将 逻辑物体 的 vectors 嵌入到 hyperbolic space？
- 自由群 F_2 的 Cayley 图是**树**，但 交换自由群的 Cayley 图是**格状**的：



(12)

- 一般来说，自由群 F_n 的 Cayley 图可以嵌入到（平面的）hyperbolic disc 上，但 F_n 的 Abelianization $F_n^{\text{Ab}} \cong \mathbb{Z}^n = \mathbb{Z} \times \ldots \mathbb{Z}$ 是一个 n-维 空间的 grid，似乎不可能嵌入到平面上
- 就算考虑 表示论 (representation theory)，所有 Abel 群的不可约复表示 都是 1-维的。F_n^{Ab} 的表示，恰好是 n 个 1-维 表示的直和。没卵用！

- 上页结论是：$\mathbb{Z}^n$ 不可能嵌入到更低维的空间内，除非使用某种 fractal 方法。然而，fractals 恰好是 神经网络 这件武器的「射程范围」之外！
- 换个想法，通过 类似 weights-sharing 的 **约束**，令 神经网络 变成 permutation invariant (= Symmetric NN)
- 这方法 必需设 activation function = polynomial
- 缺点是：约束的数量 随著 层数 增加而 指数式增长，暂时只能做 1-2 层的，每层 = 2 次多项式
- 优点是：反正我们希望 神经网络 是 sparse（减少权重空间的自由度），而这个做法同时具有逻辑 bias
- 细节我会在另一篇论文讲述

Field, JH (2010). "Derivation of the Schrödinger equation from the Hamilton-Jacobi equation in Feynman's path integral formulation of quantum mechanics". In: *European Journal of Physics*.

Nickel, Maximillian and Douwe Kiela (2017). "Poincaré embeddings for learning hierarchical representations". In: *Advances in neural information processing systems*, pp. 6338–6347.

多谢收看 ☺️