

AGI 的一些基本概念

YKY 甄景贤

Independent researcher, Hong Kong
generic.intelligence@gmail.com

May 4, 2019

Talk summary

- 1 什么是 归纳偏好？
「没有免费午餐」
- 2 神经网络 的 力量
来自什么？
- 3 Turing 机 与 逻辑 的 宇宙性
- 4 经典逻辑 AI 系统 的 基本结构

Section 1

什么是 归纳偏好？
「没有免费午餐」

机器学习 的 目的

- 机器学习 的 目的，是在某些「学习机器」的 空间 中，搜寻 符合要求的某些机器
- 例如在所有给定大小的神经网络中，搜寻符合 目标函数 的那些神经网络的 weights

AI Winter

- 一般来说，AI 的 樽颈问题 就是 搜寻空间 太大，导致 学习 太慢
- 历史上「AI 寒冬」出现的原因，是因为 基於逻辑 的学习方法，导致 搜寻空间的 组合数量爆炸，而沒有很好的 heuristic（算法窍门）

Inductive bias (归纳偏好)

- 每种学习方法都有它的 归纳偏好
 - 换言之，在 搜寻空间 里预先 划分 某些部分 是不会搜索的
 - 所以 偏好 令学习更快
 - 但如果 偏好 太强，连 答案 所在的空间也删除了
- “Throw the baby out with the water”

Section 2

神经网络 的 力量
来自什么？

神经网络的结构

- 一粒神经元 就是 一个 dot product 接著一个 非线性函数:

$$\sigma \langle \mathbf{x}, \mathbf{w} \rangle \quad (1)$$

- 这非线性函数 可以有很多种, 例如:

$$\sigma(\xi) = \frac{1}{1 + e^{-\xi}} \quad (2)$$

神经网络的结构

- 一层神经元 是 一个 矩阵 乘法:

$$\odot(W \cdot x) \quad (3)$$

- 一个神经网络 是很多 层 的函数 composition $(f \circ f)$:

$$[\odot W]^L x \quad (4)$$

神经网络 的 特性

- 神经网络 是一个有很多 参数 的 函数
- 它是 万能的 函数 近似器 [Cybenko 1989]
- 定理 的 证明 可追溯到 Weirstrauss 定理，即：任意连续函数 可以用 多项式 近似

神经网络 的威力来自「深度」

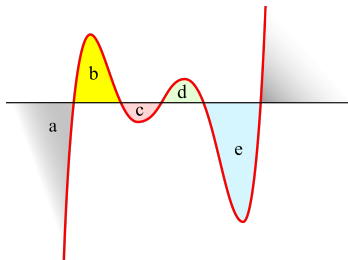
- 例如，假设 \odot 是 3 次多项式
- 每增加一层 神经网络，等如

$$(\text{多项式} \circ \text{多项式}) \quad (5)$$

- 故，总体的 多项式 次数 $= 3^L$
- 换句话说 整体次数 呈 指数式增长

神经网络 的威力来自「深度」

- 代数基本定理：多项式 次数 = 曲线 跨过 $x = 0$ 多少次



(6)

- 高维：曲面对分类空间 分割成 多少块

神经网络 的威力来自「深度」

- 这和 VC-dimension 道理一样 [Vapnik–Chervonenkis 1971]
- VC-dimension = 函数 能将 空间 分割成多少块
- 多层 神经网络 的 VC-dim 是 $O(N \log N)$ 其中 N 是 网络参数 的总个数，但证明用的是不连续的 阀函数
- 我估计 VC-dim 会是 指数增长的，但未有证明

神经网络 的威力来自「深度」

- VC-dim **指数增长** 的意义，表示 神经网络 能 代表 一个 非常复杂的 **函数家族**
- 而 神经网络 的参数个数 相对地 很少，可以在 电脑上实现

卷积神经网络的启示

- Yann LeCun 在 1989 发明了 ConvNet, 彻底改革了 机器视觉领域, 最近得了 Turing 奖



卷积神经网络的启示

- CNN 将普通 NN 的点积用卷积代替：

$$\boxed{\text{点积}} \quad \mathcal{O}(\langle \boldsymbol{x}, \boldsymbol{w} \rangle) \rightsquigarrow \mathcal{O}(f * g) \quad \boxed{\text{卷积}} \quad (7)$$

- 而卷积具有**平移不变性**，有利于视觉：

$$T_x(f) * g = T_x(f * g) \quad (8)$$

- 这是一种**归纳偏好**，令学习更快

卷积神经网络 的 启示

- 其实 视觉 需要的是 仿射 (affine) 不变性，它包括 平移、旋转、放大缩小 等
- 但 似乎 单单是 平移不变性 所带来的 学习加速，已足以令 CNN 在 2012 年超越了人类水平
- 可见，归纳偏好 在 深度学习 里 仍然是很有用的

Section 3

Turing 机 与 逻辑 的 宇宙性

有限自动机

- For example, the logic rule “love and not loved back \Rightarrow unhappy” performs the rewriting of the following sub-graph:

$$\mapsto \quad (9)$$

- This is the **state transition** $\vec{u} : \vec{x} \mapsto \vec{x}'$, which can also be regarded as the **logical inference** $\vec{u} : \vec{v} \vdash \vec{x}'$, where \vec{u} is the rewriting function or logic rule.

The problem with predicate logic

$$\forall x, y, z. \text{father}(x, y) \wedge \text{father}(y, z) \rightarrow \text{grandfather}(x, z) \quad (10)$$

- This involves **variable substitutions** which are troublesome to handle with neural networks.
(The difficulty seems to come from the cylindric-algebraic structure of predicate logic: if a formula have variables

Relation algebra

Given that:

$$\text{Father} \circ \text{Father} = \text{Grandfather} \quad (11)$$

we can deduce:

$$\text{john Father paul} \quad (12)$$

$$\text{paul Father pete} \quad (13)$$

$$\Rightarrow \text{john Father} \circ \text{Father pete} \quad (14)$$

$$\Rightarrow \text{john Grandfather pete} \quad (15)$$

via *direct* substitution of equal terms.

We're looking for developers to implement a prototype.

Thank you