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 inductive bias ?「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

什么是 inductive bias？「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，而没有很好的 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**

一粒神经元

- 一粒神经元 就是 一个 dot product 接著一个 非线性函数:

$$\sigma\langle \boldsymbol{x}, \boldsymbol{w} \rangle \tag{1}$$

- 这非线性函数 可以有很多种，例如:

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

一层神经元

- 一层神经元 是 一个 matrix multiplication：

$$\mathcal{O}(W \cdot \boldsymbol{x}) \tag{3}$$

一个神经网络

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

$$[\sigma W]^L \, \boldsymbol{x} \tag{4}$$

Neural network

- A neural network is a generic function with a large number of **parameters** called **weights**:

weight matrix for each layer

total # of layers

$$\boldsymbol{x}_{t+1} = \boldsymbol{F}(\boldsymbol{x}) = \sigma(W_1\sigma(W_2 \ldots \sigma(W_L\,\boldsymbol{x}))) \quad (5)$$

- σ is the **sigmoid** function applied *component-wise* to the vector $\boldsymbol{x}$:

$$\frac{1}{}$$

“Unreasonable” effectiveness of neural networks

- If  is replaced by polynomial, degree of the composite function increases **exponentially** as # layers increase

Optimization over logic formulas (2)

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

$$\mapsto \tag{7}$$

- 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

$$\forall x, y, z.\ \text{father}(x, y) \wedge \text{father}(y, z) \rightarrow \text{grandfather}(x \tag{8}$$

- 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:

$$\text{Father} \circ \text{Father} = \text{Grandfather} \tag{9}$$

we can deduce:

$$\text{john Father paul} \tag{10}$$

$$\text{paul Father pete} \tag{11}$$

$$\Rightarrow \text{john Father} \circ \text{Father pete} \tag{12}$$

$$\Rightarrow \text{john Grandfather pete} \tag{13}$$

via *direct* substitution of equal terms.

We're looking for developers to implement a prototype.

Thank you