**关于神经元训练模型**

**前言：**

本文旨在从各角度讨论分析，AI系统的演化，并提出神经网络模型的新思路。灵感来自于研究生课程中的Bio-inspired AI 课程，这是一门通过研究生物行为来设计AI系统演化规则和仿生机器人的课程，同样也启发了对于神经网络模型的思考。同时本文也参考了，教育学中的阈值概念threshold concept，对于神经网络模型的训练更新有指导性的思路。在对于神经网络的探索过程中，一些对于哲学的思考也可以帮助解释最终神经网络的一些不易理解的输出问题。因为本文是对于方向性的探索与论证，所以将不涉及具体的算法模型。接下来会按照这样的顺序来进行讨论：神经元模型的工作原理——根据工作原理推导总结出来人脑的认知过程及模型——通过例子证明认知模型的可行性——并通过教育学的某些概念指导训练认知模型的自训练和更新——通过哲学的角度提出自训练认知模型的解释方向——最后是关于不同神经区域的通信机制对神经网络的参考，以及部分有待思考的方向。

AI不是某一门独立的学科，其探究过程应该综合各方面，来指引AI的发展方向，可以参考很多学科的模型，如：神经模型，生物模型，社会学模型，化学模型，教育学模型等等。

**引子：**

基本的神经单元通过spike传导电信号给下一个神经单元，其中经常有stimulus的神经通路的神经元间的volt threshold相比于其他通路，会更低，也就是其被激活的门槛更低，反应更加迅速。在现实中可以对应为潜意识的反应，类似于某些脑筋急转弯的问题就是利用这样的通路来误导人的认知反馈。这一过程可以形象的称之为“联想”。，

对应到CNN模型中，算法的每一个点的weight会随着输入和输出来进行调整，这样的过程就是神经网络中寻找最快神经链路的过程。也就是根据输入得到的最近“联想”过程，通过“联想”得到对应的输出。这个输出结果并不一定是“正确”的，只是根据record得到的与输入pair的最短链路。

所以根据以上的工作原理，我们得到了一个简易的“联想”模型，而这个模型极其的强大，以至于能为任何已存在与数据库的输入和输出找到匹配。类似于case的反馈，一旦接收到某个特定的case将会给出其对应的输出，但区别在于case过于严格的输入要求，而CNN网络可以将未知的输入匹配到相对准确的输出需求。（这样的原因可能在于神经网络会自发生成一套对输入的未知分析系统，其分析的逻辑将输入简化，分割，然后将分析后得到的结果与输出匹配，这样即可完成对未知输入的解析）

本文不再对CNN模型进行解释，总之其优良特性在生物神经模型中都具备合理的理论解释，如权值共享、可训练参数少、鲁棒性强等等。对于不同的任务，使用不同的网络结构和层数，其中的原因也一定具备其生物学的解释。

**教育学的某些研究成果同样适用于对AI的学习训练：**

在教育学中，有threshold concept（阈值概念）。这个阈值概念的意思大概可以被解释为，人对未知的概念的接受过程，如同走过一条黑暗的隧道，并打开隧道另一头的门，这过程中会碰到墙壁，并具有对未知本能的恐惧，但是在最后打开大门，走出隧道的时候，才算是接受了新的概念，在这样走过隧道的过程中，需要的是类似于老师这样的角色，来引导方向的正确性，并驱逐恐惧，降低碰壁的次数。同样，在AI训练过程中，对特征点的提取，或者在更高层次的修改与调整时，需要人为的手动调试，可以大幅减少训练的时间和对数据的需求。如果对于每一层网络的理解更加具体，更会帮助这样的调整过程，当然这一过程也可以由AI进行指导。

同时阈值概念存在一个问题，一旦学习者完成了学习，那么学习者再也无法回到学习前的状态，也很难回忆起学习前所遇到的困难。这个过程好比，人脑中一旦建立了短神经链路，就很难被恢复到从前，对应到算法中，一旦权值被确定下来，就很难再回到初始状态。这样的规则所表现出的形式是人们对于某种输入的固有印象，对于这个问题的解决办法并不是消除之前的印象，而是对于同样的输入建立新的输出，在新输出的建立过程中会将新输出和固有输出进行对比，然后选择输出。但其**优点**体现在大量减少对相似问题的学习时间，并将相似的输入通过相似的链路输出，只是在最后一层增加选择项，进行对比和选择得出最后的结果。

举个通俗的例子：将百事可乐与可口可乐分别放进同样的透明玻璃杯中，只通过观察，可以得出里面装的是某种可乐。假设经过多次训练后，之前每次品尝的都是百事，突然品尝到一杯可口可乐时，会先入为主的认为这杯百事有点不一样，但不会觉得这是完全不同的饮料，因为至少从各个维度来看都是极其相似的，但被告知结果后，进行的学习修改会是在不改变之前固有的输出链路的情况下，创建一个新的相同层级的网络对比，比如从含气量或甜感度来获得不一样的输出结果，得到新的输出匹配后与之前的进行对比，最后得出结果，这样可以大量减少训练的时间。

**Reference：**

Cousin G. An introduction to threshold concepts[J]. 2006.

Meyer J H F. Threshold concepts within the disciplines[M]. Sense Publishers, 2008.

Meyer J, Land R. Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines[M]. Edinburgh: University of Edinburgh, 2003.

**模仿学习**：

人类本质确实为复读机，大家一切的行为，以及对行为动作的理解，都源于以往的经验的重复。但其中的反馈机制对对应到神经元究竟是怎样的情况，可以通过以下三篇来窥探一二。因为下面三个链接中的文章已经讲述的很清晰，如果再进行加工很可能丢失其中的细节，所以就采用了直接引用的方式。（因为比较难总结）

**镜像神经元 （mirror-neuron）**

<https://www.sciencedirect.com/topics/neuroscience/mirror-neuron>

**镜像神经系统与动作认知（The mirror neuron system and action recognition）**

<https://doi.org/10.1016/S0093-934X(03)00356-0>

MIT制作的人体行为预测系统很可能就是参考了以下文章，对镜像神经元的研究的潜在价值将在各个领域得以体现，比如自动驾驶，AI助手等。

以下是Rizzolatti教授的某篇文章中的部分：

观察其他人执行给定的运动行为，会引起观察者大脑中的运动激活，类似于观察者自己所发生的运动。计划并执行该操作。由于这种目标相似性，观察者将他人行为的目标视为自己的目标。他们不需要任何推理处理（Rizzolatti等，2001； Rizzolatti＆Sinigaglia，2008）。

有两项研究支持这种解释，其中实验范式允许猴子了解在没有视觉信息的情况下实验者所进行的运动行为的目标。在两项研究中，均记录了F5区域的神经元。在一项研究中，猴子听见了运动行为的典型声音，例如撕碎纸片或弄碎花生，却没有看到它（Kohler等，2002）。在另一项研究中，在两种情况下对猴子进行了测试：首先，他们看到实验者的手朝着物体移动以抓住它；在第二个实验中，实验者执行了相同的动作，但由于插入了不透明的屏幕，猴子只看到了它的开始部分（伸手可及的地方），而没有看到最后一个部分（抓握）（Umiltà等，2001）。）。数据显示，在两个实验中，镜像神经元在没有适当视觉信息的情况下都会放电。因此，神经元的激活反映了对他人运动行为目标的理解，而不是描述该运动行为的感觉信息。

“On an opposite cognitive side, there is a long philosophical tradition that maintains that action understanding is based on the capability of individuals to ‘read’ the mind of others, that is, to attribute a causal role to their mental states (such as beliefs and desires) in representing and executing actions. While the nature and the format of this ‘mindreading’ is controversial (Carruthers & Smith, 1996; Goldman, 2006; Hutto & Ratcliffe, 2007; Malle et al., 2001), there is no doubt that humans are endowed with this capability.”-------Action Understanding G. Rizzolatti, in Brain Mapping, 2015 <https://doi.org/10.1016/B978-0-12-397025-1.00354-7>

这一段英文比较难于理解，因为涉及到了某些未接触过的哲学概念，所以只能直接粘贴过来，但其涉及到的神经系统对自我和他人的个体的认知，是值得AI系统去思考的探索的。

**关于哲学解释**：

根据哥德尔不完备定理，作为一个低纬度的定义体系，永远无法100%的描述或窥探的高纬度现象。所以很多神经反应其实并没有实际的意义，只是将某个特定的反应与表达现象进行“绑定”后，其具备了一种definition的意义。通过哲学虚无主义得出的结论是，万物本身并没有实际意义，只是被人赋予了意义，所以其存在才具有意义。而人类所赋予的意义和理论体系是永远不可能涵盖100%的高纬度现象的，只能尝试解释某种规则的存在，也就是“道”，万物的运行法则。

这一点通过bio-inspired的小实验可以体现出来，比如某种细菌通过迷宫寻找到食物，没有感知的细菌最终会到达迷宫的终点找到食物得以繁殖。生物进化的原理在于演化更替，食物作为生存激励，让生物向前而行。根据这一点，可以制作一个简单的二维世界的规则，并让这个世界运作起来，比如在一个围棋盘上，有这样一个规则：一个白格的左右两个会生成黑格，一个黑格上下两格会生成白格。然后给棋盘一个中心奇点，让这个格子变为白格，接下来就交给时间，黑白格会铺满整个棋盘。但是当黑白格的生存条件具备竞争关系时，情况又会变得不一样。比如：白格会吃掉上下两格的黑格，而黑格会吃掉左右两格的白格。这样棋盘上的情况就变成了一个动态竞争的情况，但是最后棋盘是否会趋于稳定成为一个死局，还是一直处于动态的范围呢？由规则决定了最后棋局的dead or dynamic。

将这对应到神经网络系统时，我们可以得到思考，为什么不能让神经网络根据规则自行竞争，最后仿真得到成品的网络系统呢？当然其中存在一个问题，就是神经网络并不具备自我更替的规则，根据DNA的复制和生命的演化来看，我们完全可以赋予神经网络10%的random seed，并给与正向的激励来让其向着我们需要的方向去演化。Facebook 所制作的两个AI对话最后输出了人类无法理解的语言的沟通，这样的问题是需要避免的，但是也很容易理解，对于机器来说，人类的语音并不是最高效的共同方式。甚至于我跟我的室友长时间相处后，只需要一个眼神就能明白对方的意图，而不需要言语的沟通。但是这样的情况并不是研发者希望看到的，所以制定合适的规则进行约束是必要的。

就先写这么多了，其他涉及的方面太广，很难一次全部整合进对于神经网络的思考中，比如：神经网络的通信，神经组网的互联同步，AI系统与传统计算的辅助系统等等。本文可能很多地方无法表达到位，但是没有办法，我们的定义体系并不能完美的sync，我只能尽量去解释，并举例一些易懂的模型。我个人很希望能够参与AI事业的开发，这是类似于少年对梦想的追逐和渴望，但是很想知道相关工作的具体内容，毕竟如果工作内容与预期不相符合的话，很可能就是浪费时间而已。

Kind regards，

徐鹏展