# **大脑的层次模型假说**

刘逸川 2019.3.27

大脑中存在着很多层次。从层次的角度理解大脑，有希望构建一个统一的模型，解释从神经元层次到智能层次的各种现象。

大脑对世界的信息进行反映和处理，也就是大脑层次和世界层次产生联系的过程。这两个层次是解释大脑时最宏观的两个层次，它们的基本联系是：大脑反映了世界层次的大量信息，世界对大脑层次的反映是大脑本身和大脑活动的影响。可以看到大脑和世界是互相包含的，两个层次纠缠在一起。这种纠缠是著名的自我指涉（self-reference）的体现。自我指涉体现在大脑的各个层次上，是从层次角度解释大脑的关键。

自我指涉的本质是不同层次产生了直接联系。例如“这句话是假话”，这句话是一个层次，这句话中的“这句话”是另一个层次。这句话层次包含了“这句话”层次，“这句话”层次又包含这句话层次，两个层次紧密联系产生了自我指涉。这个例子比较特殊，事实上任何两个层次，只要存在联系，便是自我指涉的推广，例如两个层次的某种性质相似，因此产生联系而产生某种程度上的自我指涉。

自我指涉最重要的性质是自我指涉，在层次上表现为层次的包含、叠加等关系。小层次可以由于联系而构成大层次。联系不仅使小层次之间形成自我指涉，而且使小层次构成大层次。这两个作用（作为两个层次）如果是同一种联系导致的，那么联系本身又构成了自我指涉。大层次作为由层次构成的层次，必然会反映小层次的一些性质，小层次和大层次因此产生联系，这是由联系产生的联系，或者说是由自我指涉产生的自我指涉。大层次之间产生联系其实很大程度上是内部的小层次和对方的小层次产生联系，但是又不完全如此，因为大层次毕竟不只是小层次的堆积，其本身会产生新的性质，也就是涌现（emergency）现象。

自我指涉还有很多重要性质，将在下文中结合具体例子分析。接下来将分别解释神经元的连接、信息处理和认知模型三个主要层次中的自我指涉。

**神经元的连接**

大脑将信息保存在神经元的连接中，这是至少两个层次互相联系的高效保存方式。突触权重和神经元阈值作为一个层次，神经元的连接结构作为另一个层次，两个层次都存储信息并且紧密联系。没有无权重的连接，也没有无连接的权重；权重、阈值和连接的变化是互相影响的。神经元的连接结构不仅为神经元安排位置，而且其本身便包含了大量信息。计算机的这两个层次是分离的，信息主要以0、1序列保存，计算机的结构几乎只是提供存储信息的空间。因此，对于同一个场景的记录，大脑就像是拍照，而计算机就像是用语言描述，计算机很容易漏掉重要信息，因此在一些需要“直觉”的工作上远不及大脑。不过计算机可以通过软件实现这种层次联系，例如卷积神经网络和词向量取得成功的很大原因便在于体现了数据本身的结构或关系信息，这样的信息表征方式使信息更加完整和易于处理。因为现实世界的信息本身便是多层次的，很多信息只有通过信息之间的关系才能表现，有的甚至要通过关系的关系来表现。这体现了自我指涉的自我指涉性质。

赫布（Donald Hebb）提出，同一时间被激发的神经元间的联系会被强化。两个神经元作为两个层次，联系是同时激发。将时间划分出层次，两个神经元如果激发时间不一样就处于不同的时间层次，同一时间是两个时间层次形成的自我指涉。如果只是激发时间相近，那就是自我指涉的推广，相近的程度表现自我指涉的强度，会在神经元的连接中表现出来，空间层次上也是同样的道理。那么考虑层次的叠加问题，显然如果位置相近的神经元同时激发，它们的联系增强是最显著的。虽然神经元可以通过轴突和很远的神经元联系起来，但是这样相连的神经元和位置很近的神经元还是不一样的，因为前者需要消耗大量资源而一般不会被采用。这体现出层次的难以取代性，在空间层次上的不足，理论上是可以通过很多很长的轴突来弥补，但是由于种种限制难以这样弥补。这种难以取代性是非常常见的，其实通过神经元的连接记录信息的方式也可以被0、1序列取代，但是这样就要记录下大量的关系信息，处理起来也更加复杂，所以并不会这么做。

**信息处理**

和大脑保存信息的方式不同，输入的信息被转换为动作电位，但是即使是动作电位也带有结构层次的信息，通过出现在特定位置和被特定神经元发出实现。比如视觉和听觉信号出现的位置不一样，视锥细胞和视杆细胞发出的动作电位反映不同的性质。这就形成了信号的多层次表示，和通过神经元的连接保存信息是一致的。

现实世界的各种信号反映到大脑中都是动作电位，只是模式的不同，而且相应的，皮层具有通用性。对于大脑而言，不同感官的信息很大程度上是没有区别的，因为它们都能被同一训练过的皮层处理，只是在模式层次上不同，而模式层次是难以避免的划分。虽然不同模式在很大程度上是一致的，但是没有皮层能够同时处理多种模式。这当然可以理解为不同模式差别太大，但是更本质的原因是划分层次提高效率。

大脑信息处理的整个过程是将现实世界的不同类型的信息转化为在一定层次上统一的信息（动作电位和模式），然后在特定的功能区分别处理，最后再整合成对世界的感知。先在一定程度上消除不同类型信息的差别，再分别以相似而有所不同的功能区处理，最后再汇聚。因为这是对客观世界的各层次特性的充分反映，多种层次处理体现了事物的多种性质，最后联系为一个层次体现世界的整体性。

大脑中的反馈非常普遍而重要，从神经元到神经网络都出现了大量反馈。反馈的本质是一个层次对另一个层次施加影响，后一个层次又反过来影响前一个层次。而且两种影响都可能经过了很多层次后才出现。这引出了自我指涉的程度问题，如前所述，自我指涉是有强弱的。事实上，最强的自我指涉是事物自己和自己的关系，其他自我指涉都是划分出了新的层次产生的。再次以“这句话是假话”为例，“这句话”指代的是这句话，但是和这句话本身是不同层次了。指代就像分隔了一下，使“这句话”和这句话的联系不那么直接了，虽然二者最终表达的都是这句话。分成两层后自我指涉就削弱了，分成更多层的自我指涉就更弱，但是如果这些层次之间的联系紧密，如同一个层次，自我指涉仍然可以很强。以反馈为例，一个层次的信号被下一个层次接收之后，下一个层次立即反过来影响上一个层次，这种自我指涉很强。如果经过了很多层次再反过来影响，自我指涉一般就比较弱，除非经过的这些层次联系很紧密。同时反馈往往和时间有关，也就是经过很多层次需要的时间也更长，由于时间层次上的联系不紧密而导致自我指涉比较弱。这和经过很多层次导致自我指涉比较弱是一致的，再次体现了层次的叠加特性。

**认知模型**

大脑通过抽象的模型来认知世界。整个反映到大脑的世界是一个最大的模型，这个模型是抽象的，世界上所有事物都被投影到这个模型中。更小的层次也是如此，大脑对每个事物都有一个原型（prototype）。以猫为例，虽然大脑中可能存储了很多只猫，但是一定有一个猫的原型。这个原型在理解世界时不断被调用，在看到猫的图片时用这个原型进行匹配，阅读到猫的文章时，根据描述对原型进行填充。除了事物，事件也有模型，即框架（frame）和脚本（script）。用手接住投来的球，大脑不是像目前的机器人一样通过计算获得球未来的位置，而是通过经验构建出的模型进行匹配，同时根据真实信息进行修正。

模型也是有层次的，在世界模型和具体事物模型之间还有很多模型层次。比如“盒子”有一个模型，“正方体”也有一个模型，盒子模型并不是正方体模型填充的结果，虽然可以被划分到正方体模型层次下。

模型这个层次的意义是提高大脑对事物和事件的处理能力。如果没有模型，只能和记忆中过去具体的例子进行比较，那么很多本质特征可能被忽略。在现实世界中，模型这个层次是不存在的，完全是大脑归纳总结而来的。现实世界的每件事和每个物都可以作为一个层次，但是如果这样划分不利于理解和适应世界，所以这些层次通过相似性联系起来，构成更大的层次。大脑非常擅于找规律，即使只见过少量例子也能抽象出模型，这和大脑的联想能力密切相关。例如虽然很少见到老虎，但是见过猫或别的类似动物，这些动物的模型对老虎的模型的构建有很大作用。

联想的神经基础可能是不同事物有不同的神经元集合，两种事物联系起来在于各自神经元集合中的部分神经元有联系，这种联系可以反映出两种事物的紧密程度，但是如何反映出这种联系是什么呢？就像鱼和水，它们的关系自然是紧密的，但是更重要的是鱼生活在水中，这种关系的表现是通过对两个神经元集合之间连接的解读还是两个神经元集合均和表示“生活在”的神经元集合连接呢？从层次的角度来看是前者，因为鱼和水的关系密切程度是一个层次，这种关系是什么是一个层次，这两个层次是有密切联系的，因此应该由同一种模式（神经元集合之间的连接）来体现。比如不可能两个事物之间的关系是属于，而它们的神经元集合的连接很弱，神经元集合的连接是同时体现多个层次的。

我们在大脑中充满层次的前提下，通过对自我指涉性质的分析，并且将其推广，用层次联系的方式统一解释了大脑中各个层次的现象。层次之间的联系非常丰富，把握层次的关键在于层次变化后还是层次，能够用同一种方式分析。不同层次产生紧密联系使大脑能够高效保存和处理信息。层次是难以取代的，因此大脑充分利用层次的特性，使不同层次在某个目的上联系起来。大脑在认知世界时将世界划分出层次，又通过联想将不同层次联系起来。

大脑反映世界是最大的自我指涉，大脑内部的各个层次都在这个自我指涉下。因此自我指涉是自上而下普遍存在的，只是在不同层次的表现有所不同。如果深刻理解了自我指涉，那么就能够将大脑复杂的现象统一在一个模型中。本文只是提出了这种想法，通过有限的例子表现了自我指涉在大脑中的体现，其中更深刻的规律还有待进一步探索。