

# 人工神经网络:当前的进展与问题

Artificial Neural Network: The Present Advances and Problems

董 聪

(清华大学 北京 100084)

## 一、引 言

人工神经网络(Artificial Neural Network, ANN)研究热潮的再度兴起有其客观的历史背景。50年代以来,以符号机制(Symbolism)为代表的经典人工智能(Artificial Intelligence, AI)形式体系取得了巨大的成功。80年代,当人们对过去30年的成就与问题进行反思时,却不得不承认,智能系统如何从环境中自主学习的问题事实上并未很好地解决<sup>[1]</sup>。Rosenblatt(1957)的感知机(Perceptron)模型首次将自组织和自学习的思想引入智能系统<sup>[2]</sup>,这一贡献对人工神经网络30年之后的重新崛起埋下了伏笔。80年代中期,以符号机制为代表的经典人工智能的发展由辉煌转入相对停滞,原先的许多期待和承诺无法兑现。人们开始深入地探索知识发现的内在逻辑,结果发现,归纳逻辑,尤其是不完全归纳逻辑是通往知识发现的合理途径。从数理逻辑的角度讲,以演绎逻辑为基础的算法体系可以发现新的定理,却无法发现新的定律。换句话说,基于符号推理的经典人工智能形式体系在机器定理证明方面的成功和在规则提取方面的失败同属必然。Rumelhart、Hinton及Williams惊奇地发现,10余年前由Werbos发明的误差反传算法(Back Propagation error, BP)能够有效地解决多层网络中隐节点的学习问题,证明Minsky对多层网络可能不存在有效的学习方法的断言并不正确<sup>[3,4]</sup>。人工神经网络的研究在一定程度上正面回答了智能系统如何从环境中自主学习的问题,而遗传算法的新一代支持者则希望揭示学习过程在基因层次上究竟如何完成<sup>[5-8]</sup>。

本文从认识论和数理逻辑的角度出发,对人工神经网络近30年的研究与发展历程进行回顾,对一些共同关心的问题探讨,以及对该领域未来的发展趋势进行预测。

## 二、知识表达的方式及其哲学内涵

知识表达一直是经典人工智能研究的一项重要问题。从数学的角度讲,知识表达可最终归结为从一类数学空间向另一类数学空间的映射。因此,如何简化映射关系和便于计算机实现对基于机器的知识发现的研究十分重要。1900年,Hilbert猜想,

多元连续函数不能用一元连续函数的复合表示。Hilbert的猜想在1957年被著名数学家Arnold和Kolmogonov证伪。Kolmogonov证明:定义在紧致子集上的多元连续函数可用有限一元连续函数的复合精确表示。Kolmogonov的工作奠定了多层前向网络映射能力数学证明的基础。

将研究重点由存在性的精确表示转向构造性的近似描述,标志着一个新时代的开始,学术界往往忽略了这种转变的存在。事实上,这种转变对后续研究的影响极其深刻,目前引起广泛关注的许多问题均源于此。在1989年至1994年间,以Hornik为代表的一批杰出数学家的加盟使研究重点发生了根本性的转变。Hornik等人证明<sup>[9]</sup>:在一个相当宽的范围内,3层前向网络具有以任意精度逼近定义在紧致子集上的任意非线性连续或分段连续函数的能力。与Kolmogonov不同,Hornik等人的研究工作有两个突出的标志:一是对隐节点的数目不作限制;二是不断放松对隐节点函数的限定条件。Hornik等人逐渐向人们展示:在隐节点的数目不受限制的条件下,对函数逼近而言,隐节点函数的有界性要求是必要的,单调递增的限制条件并非必要。White的解释则更为直观,他得出结论说,重要的是大量神经元之间的相互连接和作用,而神经元特性的特定选择在网络逼近中的作用无关紧要。对隐节点的数目不作限制及放松对隐节点函数的限定条件,意味着生物学约束的弱化,这一点并不是生物学家们所期望的,也不是以探索智能形成机制为目标的一些AI研究者的初衷。

不论生物学家是否对Hornik等人的研究结果感兴趣,也不论一些AI研究者的初衷到底是什么,有一点是可以肯定的,即人们对Hornik等人研究结果的正确性和适用性并未产生过怀疑。将函数逼近论作为前向网络数学逼近理论基础的思想从此开始盛行。

90年代,与Hornik等人的证明工作进行展开的是不断增多的工程应用,于是,函数逼近论中从未出现过的一个概念开始在前向网络的应用中频繁地出现,这个概念就是泛化(Generalization)。前向网络泛化问题的出现使人们认识到,Hornik等人的证明曾经带给他们的信心看来多少有些脆弱。人们开始反思,函数逼近论是否足以构成前向网络数学逼近的理论依据?我们认为,经典函数逼近论事实上并不足以构成前向网络数学逼近的理论依据<sup>[10]</sup>。原因是:经典函数逼近论研究的主要是3类问题:一

是线性赋范空间  $X$  中的一个确定的元素  $x$ , 借助于  $X$  的某个给定的子集  $F$  来逼近的问题; 二是线性赋范空间  $X$  中具有某些已知性质的  $X$  的子集  $\mu$ , 借助于  $X$  的某个给定的子集  $F$  来逼近的问题; 第三类问题通常涉及寻找 (在一定意义下的) 最佳逼近集和最佳逼近方法。很明显, Hornik 等人研究的实际上是第二类问题。上述 3 类问题的共同点在于, 被逼近对象 (无论它是一个元素  $x$ , 还是一个子集  $\mu$ ) 的特性是已知的, 而在多层前向网络的应用中, 多数情况下被逼近对象的特性是未知的。众所周知, AI 中的联结机制 (Connectionism) 优于物理符号机制 (Symbolism) 的一个最主要的特征就是它所具有的学习功能, 人工神经网络之所以会再度兴起, 恰恰是因为它在一定程度上正面回答了智能系统如何从环境中自主学习这一最关键的问题。从认知的角度讲, 所谓学习, 就是对未知现象或规律的发现和归纳。因此, 经典函数逼近论的研究所基于的数学空间和多层前向网络的实际逼近过程所基于的数学空间的特性是不同的, 它们是两类不同性质的逼近问题, 具有完全不同的逼近机制。

Kolmogonov 研究的是精确表示而非逼近问题, Kolmogonov 只是改变了研究对象的表达形式, 但未改变它所具有的特性, 因此不存在泛化问题。Hornik 等人研究的是经典函数逼近论框架下的函数逼近问题, 经典函数逼近论中, 虽然新的逼近表达式中的元素和被逼近对象中的有关元素的特性不同, 但由于被逼近对象的完整特性是已知的, 逼近过程在整个定义域内实现, 无延拓现象发生, 因此也不存在所谓的泛化问题。从数学的角度讲, 前向网络的泛化问题是在被逼近对象的完整特性未知的条件下, 由有限维离散点集间的映射关系向无限维连续点集间的映射关系延拓时产生的。

上文我们解释了泛化问题的数学根源, 接下来进一步分析泛化问题的逻辑根源。

无论是联结机制还是物理符号机制, 其学习过程均表现为广义的归纳过程。联结机制以隐含的方式表示其知识, 而物理符号机制的知识表示是显现的。在归纳过程中, 不存在泛化问题, 泛化是在应用过程中出现的。无论是联结机制还是物理符号机制, 其应用过程均表现为广义的演绎过程。联结机制以隐含的方式进行推理, 而物理符号机制以显现的方式进行推理。当演绎推理的结果超出了早先用于归纳学习的原始知识的具体描述时, 泛化问题便产生了。

归纳有两种, 即完全归纳方式和不完全归纳方式。由完全归纳方式得出的结论是事实而不是猜想 (假说), 由不完全归纳方式得出的结论是猜想 (假说) 而未必是事实。以完全归纳方式得出的结论为基础的演绎推理不会出现泛化问题, 因为其演绎推理的结果必然是早先用于归纳学习的原始知识形式的某一种或某几种具体的描述, 这是一种封闭的体系, 用 AI 的术语来说, 就是这种体系只有有限的记忆功能, 而没有创造功能。泛化问题来源于以不完全归纳方式得出的结论为基础的演绎推理过程, 这是一种开放的体系, 它不仅具有有限的记忆功能, 同时还具有创造功能。网络的泛化问题是伴随着网络的创造功能而产生的, 泛化的结果既可能是

新的发现, 也可能是错误的推论。因此, 泛化问题是任何开放体系所具有的一个最基本的问题, 泛化问题并不局限于联结机制, 物理符号机制同样存在泛化问题。

Kolmogonov 和 Hornik 等人的研究不仅不涉及泛化问题, 事实上也不涉及学习问题。从数理逻辑的角度讲, 经典函数逼近论是以公理和定理为基础的演绎体系, 而从 16 世纪英国哲学家 Bacon 时代开始, 人们就已经认识到, 归纳逻辑才是通往知识发现的主要途径。当代英国哲学家 Popper 提出了关于知识发现的猜想—反驳学说, 对 Bacon 的观点做了扩展和修正<sup>[11]</sup>。从学习过程分析, 人工神经网络采用了猜想—反驳的方式, 即通过反馈环节逐步实现学习结果的优化; 从学习结果分析, 人工神经网络采用了归纳方式, 即从经验数据中提取规则。也就是说, 人工神经网络的学习过程并不具有典型的演绎特征。此外, 同其它网络的学习过程一样, 前向网络的学习也是通过反馈环节实现的。因此, 前向网络的学习过程是严格的动力学过程。所谓前向网络是根据网络的表示和连接形式, 而不是根据网络的学习过程所具有的特征来划分的, 目前流行的所谓前向网络缺乏丰富的动力特征的问题事实上并不存在。

1991 年, Brooks 题为“没有推理的智能” (Intelligence without Reason) 的论文在 AI 领域产生了广泛的影响<sup>[12]</sup>, 赞赏和批评接踵而至。Brooks 的文章使人们普遍产生了误解, 似乎人工神经网络所表现出的智能行为仅仅源于反馈。我们认为, 反馈在智能形成机制中起了重要作用, 但不是全部作用; 同时, 不以显现的方式进行推理和没有推理过程的发生是两回事。事实上, 人工神经网络不是不进行推理, 只不过推理过程未以显现的方式表达而已。原因很简单, 没有推理就不会出现有目的的行动, 没有行动就无法产生反馈信号, 而没有反馈信号的引导, 所谓的自组织行为便无法完成。从这个意义上讲, 我们认为 Nillson 对 Brooks 工作的批评准确而恰当。Nillson 在评价 Brooks 的工作时非常巧妙地引用了 Anden 的诗句: “不会推理的将在行动时消亡, 不会行动的则因此而消亡”。

人工神经网络提供了一种揭示智能和了解人脑工作方式的合理途径, 但两者之间并不存在一一对应的关系, 要真正揭示智能的形成机制和了解人脑的工作原理, 深入研究神经生物学的基本规律并在构建相应的人工模型时考虑生物学的基本约束是十分必要的<sup>[13,14]</sup>。

### 三、新的启示与猜测

把目光从数学家虚拟的背景中移开, 再次投向真实的生物系统时, 我们惊奇地发现, 在神经网络的研究中, 生物学家们正在进行的工作比人们想象的更具有启发性和创造性。

Shatz 对发育中的大脑进行了深入的研究, 他发现成熟的神经连接形式的出现在于轴突有选择地收缩及长出新的分枝的重新安排, 轴突在其靶结构 (靶位) 内伸到不同地址 (靶址), 然后再通过竞争逐

步消除初始选址的错误<sup>[15]</sup>。视觉系统中的靶位对应于人工神经网络中的隐层，而靶位内的靶址对应于隐层中隐节点的位置。Shatz 研究的是视网膜节细胞的轴突在外侧膝状核内发育及成型的过程，同样的过程也发生在外侧膝状核内神经元的轴突在初级视觉皮层（亦称 V1 区）内的发育成型中<sup>[14]</sup>。

我们将 Shatz 等人的发现以计算机模拟程序的方式再现出来，提出了多层前向网络中跨层连接的概念及网络拓扑结构压缩的通用算法<sup>[16]</sup>。对于著名的 2~7 位 Parity Mating 问题，我们的仿真结果表明，采用通用算法压缩后的网络结构为 N-N-1。也就是说，当学习样本以指数方式增长时，压缩后的网络结构，其隐节点的增长方式是线性的，人们原先担心的 NP hard 问题并未出现。我们还发现，最大限度地压缩网络的拓扑结构是实现网络泛化功能的有效途径。

重新审视一下如式 (1) 表示的含有 n 个输入、1 个输出、M 个隐节点的 3 层前向网络

$$f(X) \sim \sum_{m=1}^M \beta_m h_m(Z_m) = \sum_{m=1}^M \beta_m h_m\left(\sum_{i=1}^n a_{im} x_i\right) \quad (1)$$

如果把隐层的节点输出看作是经过变换后的广义坐标的话，则 3 层前向网络对离散点集的最佳逼近，其实质就是选择最合适的广义坐标  $h_m(Z_m)$ 。从式 (1) 可以看出，广义坐标的形成依赖于输入层到隐层的连接权矩阵  $a_{im}$  及隐层节点函数  $h_m$  的特定选择。 $a_{im}$  是无级可调的，因此可以得出结论，广义坐标的特性本质上主要取决于隐层节点函数的特定选择<sup>[16]</sup>。

网络实际学习过程中，学习样本是有限的，因此，网络中有效隐节点的数目存在上限 ( $\square N$ )<sup>[10,17]</sup>。换句话说，不可能通过无限扩充隐节点数目的方式改进逼近效果。这意味着由 White 开始的，目前甚为流行的，认为神经元特性的特定选择在网络逼近中的作用不重要的观点事实上并不正确。

70 年代以来，现代生物学的研究成果进一步揭示，动物的学习和记忆过程不仅表现为突触的变化，而且突出地表现为神经细胞内部构造的变化，即人工神经网络中节点函数的特化。生物系统发育过程中神经细胞内部构造的变化已被哈佛医学院 Selkoe 的研究结果所证实<sup>[18]</sup>。

研究智能的形成机制，考虑生物学的基本约束是十分必要的。除此之外，许多人仅把前向网络作为一种合适的逼近工具来研究和利用，这种工具主义的思潮目前在工程界甚为流行。

从工具主义的角度讲，前向网络的学习过程更类似于多元非线性回归分析。其中，投影寻踪算法 (Projection Pursuit Algorithm, PPA) 中的一些思想值得借鉴。投影寻踪算法中的岭函数、岭函数总数、岭函数自变量的系数、岭函数权重和 3 层前向网络中的隐节点函数、隐节点总数、输入层至隐层的连接权系数和隐层至输出层的连接权系数之间存在一一对应的关系，在投影寻踪算法中，岭函数和岭函数总数自身也是学习的对象。

上文给出的网络拓扑结构压缩算法实际上就

是一种隐节点总数的学习算法，而以径向基函数 (Radical Functions)，特别是以子波函数 (Wavelets Functions) 作为隐节点函数的研究，也证明了隐层节点函数特性的特定选择对网络的逼近效果具有十分关键的作用。在工具主义思潮的影响下，我们给出了 BP 算法的广义描述，证明基于广义 BP 算法的任何全样本前向网络权值修改方式都是收敛的<sup>[19]</sup>。除此之外，我们还给出了前向网络学习的多种方式，其中，按隐节点顺序选择待修改权值的方式其作用和投影寻踪类似。

从工具主义的角度出发，我们发现可变的隐节点函数特性对改进网络的逼近精度，特别是改进网络的泛化能力十分有利。Selkoe 的实验研究则证实，神经细胞内部构造在生物系统发育过程中确实发生了变化。这使我们进而推测，在生物系统中，存在的必有其合理的地方，而合理的则更有可能存在。也恰是“合理的则更有可能存在”这一信念确立了我们审视人工神经网络全局优化问题有关研究工作的基本视角。

#### 四、全局优化方式的哲学思考

人工神经网络的研究中，有 2 种随机优化算法非常流行，它们分别是 Metropolis 提出的模拟退火算法 (Simulated Annealing, SA)<sup>[20]</sup> 和 Holland 提出的遗传算法 (Genetic Algorithm, GA)<sup>[5]</sup>。SA 模拟的是金属材料加温后的退火过程，GA 模拟的是生物的自然进化过程。从模拟对象看，两者几乎没有任何共同之处，但从算法本身的形式结构看，两者则极其相象。SA 借助于产生函数 (Generation Function) 确保状态空间各点的概率可达性，通过接受算子 (Acceptation Operator) 模拟平衡态，通过接受算子的定向变化保证算法迭代进程的方向性；而 GA 则通过交换算子 (Crossover Operator) 和突变算子 (Mutation Operator) 的协同作用确保状态空间各点的概率可达性，通过选择算子 (Selection Operator) 的作用保证算法迭代进程的方向性。

从数学的角度讲，实现全局优化的随机算法可分为两大类：一类是通过遍历搜索的方式，如 SA 和 GA 等；另一类是通过定向推进的方式，如广义遗传算法等<sup>[6]</sup>。

退火是金属材料热处理的一种方法，其作用是使材料的成份更均匀，细化晶粒，增加韧性，因此，退火本身不存在全局最优化问题<sup>[8]</sup>。SA 将组合优化问题中的状态定义为相格，将各状态的函数值定义为相格所对应的内能，将每次抽样所获得的微粒 (微观状态) 按照 Boltzmann 统计分布进行取舍，并将留下的微粒向对应相格中投放，虚拟温度 (Boltzmann 常数与温度归并为一体) 则定义为系统中所有微粒所含内能的平均值。虚拟温度的定义方式和 Boltzmann 统计分布的引入，其目的就是模拟热平衡状态的形成过程，然后，通过控制虚拟温度的逐步降低，进一步模拟热平衡状态的定向转移过程。因此，SA 希望模拟的实际上是热平衡状态的形成及定向转移过程。

从热力学的角度讲，热平衡状态的定向转移只

是使系统各微观状态的內能所服从的随机分布的形态(概率密度函数的形态)发生定向变化,并不能产生新的状态。换句话说,SA拟解决的是大量抽样中系统平均內能(或峰值状态)的最小化问题,并不是单个状态的最小化问题,而后者才是全局优化方法真正需要解决的关键问题。除此之外,SA在具体操作上也存在一些问题。

从统计物理学的角度讲,所谓热平衡状态,就是—定温度下所有可能微观状态中出现几率最大的状态,因此,模拟热平衡状态需要两个条件:(1)所有可能的微观状态是已知的;(2)采用 Boltzmann 统计分布作为接受算子的前提是所有可能微观状态以等概率产生。如果说第二个条件只涉及技术难度的话,则第一个条件事实上无法满足。原因很简单,如果所有可能的微观状态是已知的,则最小值状态自然也是已知的,无进一步求解最小值之必要。

SA 经改造之后可用来解决系统中单个状态的最大或最小化问题,此时算法成败的关键在于,如何选择产生函数使状态空间各点的概率可达性得到保证,目前普遍关注的虚拟温度下降速率并不是问题的关键。可以证明,只要状态空间各点的概率可达性得到保证,采用确定性的接受算子保证算法迭代进程的方向性,系统中单个状态的最大或最小化问题便可直接解决,模拟热平衡状态的形成及定向转移过程并无必要。

同样,自然进化过程也不存在全局最优化问题。Jacob 认为,进化就像是—个修补匠,它只能从当时所能得到的材料中,有选择地进行调整(使生物体适应环境),其能力受到以前的决定所限制。不仅如此,每个特定的能力看起来都—个—个优美的设计,但实际上,它也许不能同更大—些的系统很好地整合在—起,并且可能同系统的整体布局不相匹配。我们认为,生物系统的演化从根本上说是动态的,对于每一个偶然的形态或环境上的变化,生物体皆用已经进化出来的一系列特征去应付生存考验,在此之前并没有—个经过内省的有意识的筹划。换句话说,进化的产物都是分阶段局部优化的结果,我们不可能从单纯模仿进化过程发现解决全局优化问题的诀窍。GA 成败的关键在于,如何通过交换算子和突变算子的协同作用,确保状态空间各点的概率可达性。

关于 GA 和 SA 全局优化能力的数学证明是—项有益的研究课题<sup>[5-8,20]</sup>,但本文则更关心它的哲学蕴涵。众所周知,热平衡态的形成是—个熵增过程,其极限是著名的 Helmholtz“热寂”,而生命的进化历程则具有明显的远离热平衡态的自组织倾向。人们不禁会问,人类智慧是随机脉动的产物还是定向进化的产物?如果人类智慧是随机脉动的产物的话,由什么因素来决定随机脉动的平均走向?如果人类智慧是定向进化的产物的话,又由什么因素来决定定向进化的方向?经典 GA 和 SA 的设计者相信人类智慧是随机脉动的产物,选择最终决定了随机脉动的平均走向。我们则相信人类智慧是定向进化的产物,反馈和选择的自然合成最终决定了定向进化的方向。选择的作用相当于能量的定向输入,它使自然进化过程远离热平衡态,并向自组织的方向发

展。因此,成功地模拟热平衡状态的定向转移过程才是全局优化算法的核心所在。

我们认为,—种算法要想具备实现全局最优化的功能,它只需满足两个条件:(1)它具有实现局部最优化的能力;(2)它具有从—个局部最优状态向—个更好的局部最优状态定向转移的能力。正是在这样—种思想的指导下,我们创立了广义遗传算法 (Generalized Genetic Algorithm, GGA),并解决了多层前向网络的最简拓扑构造问题和全局最优逼近问题等<sup>[16]</sup>。GGA 和前向网络最简拓扑构造算法均摒弃了包括经典 GA 和 SA 在内的随机优化方法普遍采用的遍历搜索策略,转而采用定向演化模式。定向演化模式不要求算法的迭代过程是—个遍历的 Markov 过程。遍历性要求状态空间的各点是互通的,而定向演化模式只要求状态空间的各点是单向可达的(事实上只要求最大或最小值状态的单向可达性),后者只是前者的必要条件。换句话说,定向演化模式的实现条件宽松得多。

## 五、未来发展之预测

以人工神经网络为代表的智能计算在未来 10 年的研究重点将越来越倾向于实用,突出表现在以下—些方面:

1. 数学证明将结合具体的算法有针对性地进行,而不是将数学证明游离于算法之外独立进行。

2. 人们开始认识到,揭示大脑智能之谜的道路比原先设想的要漫长得多,大脑的—些特性,如容错性、鲁棒性等有其内在的、尚未被完全揭示和深入了解的生物机制,并不必然为人工神经网络所拥有。算法的可实证性会受到越来越多的重视,因此,客观而具有针对性的量化描述将逐步取代类比和充满隐喻的语言。

3. 个性化的倾向越来越浓,目的性变得日益明确。—方面,工具主义者不再将自己禁锢于生物学约束的囚笼之中,而仅将生物学的发现看作是开拓视野的—种途径;另—方面,以探索智能形成机制为目标的 AI 研究者会越来越强调生物学约束的重要性,新的生物学发现会越来越多地作为原始素材被用于构建新的网络模型,而不是一味地被用来证明旧有模型的合理性。

### 参考文献

- [1]董聪,酆正能,夏人伟,何庆芝. 多层前向网络研究进展及若干问题. 力学进展,1995,25(2):186-196
- [2]Rosenblatt F. Principles of Neurodynamics: Perceptrons and The Theory of Brain Mechanisms. New York, 1962
- [3]Minsky M, Papert S. Perceptron. MIT Press, 1969
- [4]Rumelhart D E, Hinton G E, Williams R J. Learning Representation by Backpropagation Errors. Nature, 1986, 323 (6188):533-536
- [5]Holland J H. Adaptation in Natural and Artificial Systems. MIT Press, 1992
- [6]董聪. 广义遗传算法. 大自然探索,1998,17(1):33-37
- [7]董聪,郭晓华. 广义遗传算法的数学结构. 中国科学基金,1998,(4):
- [8]董聪,郭晓华. 广义遗传算法的逻辑结构及全局收敛性的证明. 计算机科学,1998,25(6):

## 美国科学研究与研究生培养一体化给我国的启示

### Enlightenment of the Combination of US Scientific Research and Post-Graduate Education

肖广岭

(清华大学人文社会科学学院科技与研究研究所, 副教授 北京 100084)

进入 20 世纪以来, 特别是第二次世界大战结束以来, 美国不仅在科学上越来越领先于其他主要发达国家(德国、法国、英国、日本), 而且在研究生培养方面, 特别是博士生培养, 不仅在规模上, 而且在质量上也大大高于其他主要发达国家<sup>[1]</sup>。在一个既没有国家科学技术部, 又没有国家教育部, 也没有国家科学院(美国科学院类似中国工程院, 只是荣誉性机构)的国度, 科学研究和研究生培养却能取得如此巨大的成就, 其条件、体制、机制和政策值得研究, 特别是其科学研究与研究生培养一体化能给我们很多启示。

#### 一、美国科技体制和研究生培养及其特点

美国的总体科技布局大体分为 4 个部分: 一是大学, 被定义为科学之家, 主要从事基础研究和应用研究, 1990 年的 R & D(研究与发展) 经费为 220 亿美元(包括大学管理的联邦实验室, 其经费为 50 亿美元); 二是工业界, 主要进行发展研究, 1990 年的 R & D 经费为 1 080 亿美元; 三是政府的科研机构, 主要是联邦实验室, 主要从事指令性(mandated)研究, 也有一些相关的基础研究和应用研究, 1990 年的 R & D 经费为 160 亿美元, 但在其大约 700 个联邦实验室中有 500 个是在大学的校园里并由大学管理; 四是独立的、非营利的科研机构, 从事公益性研究和其他各种研究, 1990 年经费为 50 亿美元<sup>[2]</sup>。

与其他发达国家相比, 美国的显著特点是基础研究及相关的应用研究(大学的这些研究常常被称

为学科性的 disciplinary 研究, 与指令性的 mandated 研究相对应)主要由大学来承担。而其他国家, 如德国的此类研究除了由大学来承担外, 普朗克研究院也进行此类研究, 并且规模较大; 法国此类研究除了大学承担外, 大学之外的国家科学研究中心也进行此类研究, 并且规模也较大。

在研究生培养方面, 美国的显著特点是教学和科研并重、科学研究和研究生培养一体化, 并且大学研究生的规模巨大。以博士生为例, 美国的博士生要花大约两年的时间修完课程, 还要经过相关考试; 而英国、法国和德国没有严格的课程要求, 大学本科毕业后跟教授做几年(一般为 3 年)的研究工作, 完成博士论文就能获博士学位。在科学研究和研究生培养一体化方面, 美国的优势在于, 大学教授掌握更多的研究经费, 而研究生是课题组的主要成员。与美国相比, 其他主要发达国家的大学及其教授的研究经费要少得多。例如, 德国和法国的同类研究经费有相当大的部分流到了大学以外的研究院或研究中心, 这就使得大学及教授的研究经费较少。尽管这两个国家的大学与大学之外研究机构实行一些联合培养研究生的方案, 但与美国的科学研究与研究生培养一体化相比还相差很远。

在研究生规模方面, 美国很多研究型大学的研究生占在校生总数的 1/2 以上, 例如, 1991 年哥伦比亚大学的研究生比例为 66.2%, 芝加哥大学为 62.9%, 哈佛大学为 61.8%, 等等。美国很多研究型大学研究生的数量超过 1 万人, 例如, 1991 年哥伦比亚大学研究生人数 11 389 人, 哈佛大学为 11 399 人, 密执根大学为 13 102 人, 等等<sup>[2]</sup>。美国大学如此之大的研究生规模远远大于其他主要发达国家的

- [9] Honik K. Approximation Capabilities of Multilayer Feedforward Networks. Neural Networks, 1991, (4): 551 - 557
- [10] 董聪. 多层前向网络的逼近与泛化机制. 控制与决策, 1998, 13(增刊): 413 - 417
- [11] Popper K. Conjectures and Refutations: The Growth of Scientific Knowledge. Harper & Row, Publishers, New York and Evanston, 1968
- [12] Brooks R. Intelligence without Reason. IJCAI - 91, 1991
- [13] Chuchland P S, Sejnowski T J. The Computational Brain. MIT Press, 1992
- [14] 董聪. 大脑、知觉模型和计算机模拟. 科技导报, 1997(7): 7 - 10.

- [15] Shatz C J. 发育中的大脑. 科学, 1993, (1): 11 - 19
  - [16] 董聪. 多层前向网络的逼近机理与拓扑结构学习方法. 通讯学报, 1998, 19(3): 29 - 34
  - [17] 董聪. 前向网络全局最优优化问题研究. 中国科学基金, 1997, (1): 23 - 29
  - [18] Selkoe D J. 大脑衰老, 智力减退. 科学, 1993, (1): 69 - 76
  - [19] 董聪, 刘西拉. 广义 BP 算法及网络容错性和泛化能力的研究. 控制与决策, 1998, 13(2): 120 - 124
  - [20] Kirkpatrick S. Optimization by Simulated Annealing: Quantitative Studies. J. Statis. Phys., 1984, 34: 975 - 986
- (责任编辑 肖庆山)