中国科学: 信息科学 2021 年 第 51 卷 第 12 期: 1967–1978
SCIENTIA SINICA Informationis 中国科学院学部 科学与技术前沿论坛
《中国科学》杂志社 SCIENCE CHINA PRESS
观点与争鸣

人工智能的 10 个重大数理基础问题

徐宗本[1,2,3]

1. 西安交通大学西安数学与数学技术研究院, 西安 710049
2. 琶洲实验室, 广州 510335
3. 鹏城实验室, 深圳 518055
E-mail: zbxu@mail.xjtu.edu.cn

收稿日期: 2021–07–04; 接受日期: 2021–09–30; 网络出版日期: 2021–12–09

国家自然科学基金重大项目 (批准号: 11690011, U1811461) 资助

摘要 本文提出并阐述人工智能研究与应用中亟待解决的 10 个重大数理基础问题: (1) 大数据的统计学基础; (2) 大数据计算的基础算法; (3) 数据空间的结构与特性; (4) 深度学习的数学机理; (5) 非正规约束下的最优输运; (6) 如何学习学习方法论; (7) 如何突破机器学习的先验假设; (8) 机器学习的自动化; (9) 知识推理与数据学习的融合; (10) 智能寻优与人工智能芯片问题.

关键词 人工智能, 数理基础, 统计学, 大数据算法, 数据空间, 深度学习, 最优传输, 学习学习方法论, 机器学习假设, 机器学习自动化, AI 芯片

1 引言

作为新一代信息技术的代表, 人工智能 (artificial intelligence, AI) 已经广泛应用于科学、社会、经济、管理的方方面面, 已经成为创新驱动发展的核心驱动力之一. 然而, 就其技术发展而言, 人工智能还只是突破了从 “不可用” 到 “可以用” 的技术拐点, 从 “可以用” 到 “很好用”“用得好” 还存在诸多技术瓶颈, 正呼唤重大技术变革.

技术变革的先导是理论创新, 即基础研究. 它是指对事物本质和规律的科学化探寻和揭示, 是启发、促动技术变革的激发源和理论依据. 理论创新既应包括对原有理论体系或框架的新突破、对原有理论和方法的新修正和新发展, 也包括对理论禁区和未知领域的新探索. 本文主要关注人工智能技术发展当前亟待解决的重大数理基础问题.

为什么要特别关注 AI 的数理基础问题呢? 这是因为当前人工智能技术和发展主要是靠 “算例、算法、算力” 所驱动的, 其基础是数据, 其核心是算法, 这二者都深刻地以数学为基础. 数学主要提供所研究问题的形式化手段、模型化工具和科学化语言. 没有形式化就没有程式化和计算机化, 没有模

引用格式: 徐宗本. 人工智能的 10 个重大数理基础问题. 中国科学: 信息科学, 2021, 51: 1967–1978, doi: 10.1360/SSI-2021-0254
Xu Z B. Ten fundamental problems for artificial intelligence: mathematical and physical aspects (in Chinese). Sci Sin Inform, 2021, 51: 1967–1978, doi: 10.1360/SSI-2021-0254

型化就没有定量化和知识化, 没有科学化就没有系统化和现代化. 所以, 数学在科学技术中具有独特的作用和价值. 对人工智能而言, 数学不仅仅是工具, 还是技术内涵本身, 而且常常也是最能体现本质、原始创新的部分. 美国国家研究委员会的报告 [1] 中所说的 “进入高技术时代, 我们也就进入了数学技术的时代”, 尼克松科学顾问艾德华 · 大卫 (David Edwards) 所说的 “很少人认识到当今如此被广泛称颂的高技术在本质上是一种数学技术” 大概都指这一道理.

2 亟待解决的 10 个重大数理基础问题

我们提出当前 AI 亟待解决的 10 个重大数理基础问题, 如下所述.

(1) 大数据的统计学基础

当下人工智能的主流技术 (如深度学习) 是以对大数据的加工处理为基础的, 它的模型、分析、计算基础都根置于统计学.

统计学一直被认为是主导和引导人们分析和利用数据的学科. 传统上, 它根据问题需要, 先通过抽样调查获得数据, 然后对数据进行建模、分析获得结论, 最后对结论进行检验. 所以, 传统统计学是以抽样数据为研究对象的, 遵循了 “先问题, 后数据” 的模式和 “数据 $\rightarrow$ 模型 $\rightarrow$ 分析 $\rightarrow$ 检验” 的统计学流程. 当今 “拥有大数据是时代特征、解读大数据是时代任务、应用大数据是时代机遇” [2], 呼唤了 “先数据, 后问题” 的新模式. 这一新模式从根本上改变了传统统计学的研究对象和研究方法, 更是动摇了传统统计学的基础. 要解决人工智能的基础问题, 就必需首先解决大数据统计学基础问题 [3,4].

熟知, 统计学是建立在概率论, 特别是像大数定律、中心极限定理、正态分布理论等这样一些基本数学原理基础上的. 这些基本原理大都是在独立同分布 (iid) 样本和观测变量个数 p 远少于数据量 n (即统计学常说的 $p \ll n$) 的假设下被证明的. iid 假设意味着样本须来自同一总体而且样本独立抽样, $p \ll n$ 假设指 “问题本身并不复杂而积累的经验 (观测) 不少” (用线性方程组来理解, 相当于 “方程的个数大于未知量个数”). 这两条假设是如此基本和影响深远, 以至于统计学中的许多原理都以此为规约. 例如, 一个观测模型的误差与系统内部变量无关, 或者说误差和结构不相关 (外生性假设), 这是统计学一直以来的公设. 很显然, 所有这些假设在大数据情形都常常不满足, 甚至会被彻底破坏. 例如, 自然记录/收集的数据既不可能仅来自于同一总体, 也不可能保证彼此相互独立; 像图像这样具有任意高分辨率 (像素个数 p) 的数据, 任何图像集合 (个数为 n) 都不可能满足 $p \ll n$, 而已有大量研究说明, 当 $p \ll n$ 破坏之后, 就必然会出现 “伪相关” 和 “内生性” (incidental endogeneity) 等伴生问题 [5]. 为了能将 AI 模型与分析置于坚实的大数据分析框架中, 显然我们需要在各种非 iid、非 $p \ll n$ 条件下去重建大数定律、中心极限定理等概率论工具, 我们也需要在真实的大数据条件下去建立各种估计的大样本性质. 这是建立可信、可解释人工智能的必备条件.

统计学与人工智能有很强的关联性但又有显著区别. 统计学使用专有的随机变量或分布函数法去建模数据, 但假设空间多限于线性或广义线性模型, 此时模型参数具有解析形式, 可通过大样本分析去建立相应推断的可解释性理论. 人工智能不提供对数据的建模, 但使用像深度神经网络这样高度复杂的函数去建模, 即使用高度复杂的假设空间, 具有应用的普适性, 但由于模型参数难有解析形式, 也必然带来推断的不可解释性. 这两种建模方法显然有着各自不同的优势与劣势, 能否将这二者融合? 如能, 如何融合? “巧用简单模型、局部拼接整体、逻辑与非逻辑混合、内核 + 边界、图网络” 等都是值得尝试的路径. 所有建模都必须在表示的广泛性和统计推断的易实现性或可解释性之间取得平衡, 这是所有方法的瓶颈. 另外, 大数据分析方法、大数据抽样理论、大数据假设检验等也都是亟需建立的统计学新理论.

(2) 大数据计算的基础算法

人工智能算法本质上是大数据分析处理算法, 主要解决大数据分析与处理技术底层依赖的相关数学模型、分析原理与计算方法等问题. 它是人工智能技术与应用的基础算法与理论支撑, 是数学与计算科学深度融合的一个新领域.

大数据分析处理的核心是通过计算对大数据进行加工处理和从中萃取有用信息. 它通常包含查询、比对、排序、化简等数据处理任务和聚类、分类、回归、降维、相关性分析等数据分析任务. 无论是数据处理还是数据分析, 它们都是通过合适的计算机算法实现的. 这些算法在 AI 中被称为核心算法. 核心算法的核心步骤通常要求在大数据环境下去解一些基本的数学问题, 求这些基本数学问题的算法被称为大数据计算基础算法. 当前, 人工智能应用的主要障碍之一是, 对真正的大数据, 大部分已知的核心算法和基础算法失效 (要么不能用, 要么算不出满意结果), 例如, 还没有一个好的算法能对超过 TB 级的数据进行直接聚类 (参见文献 [4]).

缺乏这样的大数据算法之根本原因在于传统计算理论, 以及基于传统计算理论的算法设计与分析方法学在大数据环境下失效. 对任何一个大数据分析和处理问题, 设计出一个超低复杂性的算法都不是简单的事. 正因为如此, 美国国家科学院/全国研究理事会在其发表的报告 [6] 中, 将在大数据环境下求解如下 7 个数学问题的问题称为 “7 个巨人问题”, 并认为是重大挑战:

- 基本统计 (basic statistics);
- 广义 N- 体问题 (generalized N-body problem);
- 图计算问题 (graph-theoretic computation);
- 线代数计算 (linear algebraic computation);
- 最优化 (optimization);
- 积分 (integration);
- 比对问题 (alignment problem).

而他们所列出的大数据环境包括:

- 流环境: 数据以 “流” 的方式给出;
- 磁盘环境: 数据存储在计算外设的磁盘;
- 分布式环境: 数据存储在不同机器或边缘端;
- 多线程环境: 数据在多处理器和共享 RAM 的环境中存储.

在大数据环境下如何求解这 7 个巨人问题, 是大数据计算所面临的核心挑战. 值得注意的是, 在通常单机环境下, 求解这 7 个巨人问题都有非常成熟的算法 (可在常用的数学算法库中调用). 由此可见, 大数据对各学科的冲击是如此之基础和普遍. 大数据基础算法研究本质上受大数据计算理论的限制. 人们期望在超低复杂性 (例如至少在线性复杂性及以下) 水平上寻找解决问题的算法. 然而, 当我们准备放弃 “多项式复杂的算法是一个好算法” 这样的传统观念时, 猛然发现: 未来的路在何方?

(3) 数据空间的结构与特性

我们所处的世界由人类社会、物理空间 (这二者常统称为现实世界) 和信息空间 (称为虚拟世界) 构成 (文献 [3]). 人类社会的构成元素是人, 物理空间的构成元素是原子/分子, 而信息空间的构成元素是数据, 所以信息空间亦称为数据空间, 是由数字化现实世界所形成的数据之全体. 人工智能作用在数据空间是利用数据空间的方法认知和操控现实世界的技术.

从这个意义上, 数据空间理应是人工智能 (或更一般地, 数据科学) 最基本的认知对象. 数据空间 (或它的特定子空间) 所具有的特征、结构、运算、特性等对于解译和应用数据显然具有本质的重要性. 譬如, 数据科学面临的首要任务之一是, 如何对自然产生的图像、视频、文本、网页等异构数据进行存

储处理. 由于这些数据并不能用关系数据库这样传统的记录方式去记录, 它们常被称为是非结构化的. 我们知道, 每一类 (或每一个) 数据都有着它自己特定的记录方式, 如彩色图像用 R-G-B 这 3 个像素矩阵来表示, 可见它并不是完全无结构的 (无结构就无记录!), 所谓非结构化本质上不是说它们无结构, 而是它们的结构不统一、不规整或者相异 (如图像可能具有不同的分辨率, 也可以是从不同谱段采集的, 既有图像又有文本等). 要储存这样的非结构化数据并便于处理, 唯一可能途径是将这些非结构化数据进一步形式化, 或称 "结构化", 即在某种更加统一、更加抽象的数学结构下, 重新表达这些所有类型数据, 并基于这样的形式化去存储和处理. 这样的过程即是非结构化数据的结构化. 只要有存储, 就必然要结构化. 结构化的本质是寻求数据的数学表示, 而关键是设置一个最小的公共维度, 使其在这个维度下, 所有类型数据在数学化空间中都能得到表达 (当然, 对每一类而言, 可能会有冗余). 要找到这样的最小公共维度, 显然依赖于知晓每一类型数据的最小表示长度. 以图像为例, 我们希望知道: 图像怎样才能够最简约地被表示? 图像放到一起能互相表示吗? 图像空间有维数吗? 如有, 是多少? 等. 把不同分辨率的图像放到一起可构成一个类似函数空间的无限维空间, 这个空间内的图像可以认为是超高分辨率或无穷分辨率的, 这一空间不仅为存储不同分辨率图像提供框架, 也为理解图像分辨率的极限行为提供理论基础. 问题是: 这样的无穷维图像空间有什么特别性质? 它对超高分辨率图像会带来什么新的洞察? 澄清这样的图像空间 (类似地, 文本空间等) 整体性质, 是彻底解决非结构化数据存储的出路所在. 在这样的探索中, 产生新的、更为有效的 AI 技术是自然不过的事.

除像上述这样需要对数据空间的某些子空间 (如图像空间) 性质展开探究之外, 我们也期望对各种数据子空间的数学结构与性质展开研究. 严格地说, 当我们使用数据空间、图像空间、文本空间这样的术语时, 这里 "空间" 往往仅指 "集合", 并没有指它们已经构成数学意义上的 "空间", 因为在其中我们并没有赋予它们特定的 "运算" 和 "拓扑". 一个熟知的事实是, 当一个对象集合被赋予某种数学结构 (运算 + 拓扑) 后可成为数学意义下的空间; 一个数学意义下的空间内部元素可以按照特定规律去运算, 也能够使用一些特定工具去分析. 所以, 对一类对象 (如图像), 只有把它放在对应的数学空间中去考察, 才能有望得到规范化、严格化的分析, 从而获得更为本质的认知. 于是, 一个自然的问题是: 对常见的这些数据空间, 能不能赋予某种数学结构使它们成为数学上的空间呢? 如能, 它们又会成为什么样的数学空间? 是内积空间、赋范空间, 还是拓扑空间 (请注意, 不同的数学空间提供的分析工具是有差别的)? 应该赋予什么样的数学结构才最自然、最合理、最有利于数据分析? 让我们仍以图像空间为例说得更具体一些: 我们能不能通过赋以缩放、卷积、平移 + 旋转等操作或运算, 并选取图像差异性的一种度量, 如欧氏距离、KL 散度、Wasserstein 距离等, 使图像空间成为数学意义下的空间? 如能, 怎样的选择和搭配才能使所建立起来的空间更有利于图像分析?

除数学空间这样的分析工具外, 数据空间的代数结构也希望得到研究. 研究数据空间的根本目的是, 为人工智能技术寻找新的突破口, 为更加有效的数据分析与处理提供新框架、新工具、新方法和新技术. 只要是有利于这一目标的任何研究都应受到鼓励.

(4) 深度学习的数学机理

当代人工智能的主流技术是以深度学习为代表的. 深度学习的巨大成功极大提升了它作为普适 AI 技术的主导地位, 但另一方面, 也唤起人们对深度学习本质局限性和 "后深度学习时代" 的思考. 深度学习的独特优势是, 对任意复杂数据都有强的建模能力, 只要训练数据足够, 就一定可学习、可应用, 从而能提供普适的 AI 解决方案. 但它的致命缺陷是, 网络结构难设计、结果不具可解释性、易受欺骗等.

为什么深度学习具有这些独特的优势, 别的方法就不具备吗? 为什么它有这些致命缺陷, 它们就不能被克服吗? 理性而严格地回答这些问题, 全面认识深度学习和思考后深度学习时代 AI 的发展, 都

是核心而紧迫的问题.

定量刻画深度学习的构 – 效关系是首要的数学原理问题. 假设

$$y = \mathcal{N}(x) = f_k(f_{k-1}(f_{k-2}(\cdots(f_1(x))\cdots))),$$

$$f_i(x) = G_i\left(W_i^{\mathrm{T}}x + b_i\right), \quad i = 1, 2, \ldots, k,\ W_i \in \mathbb{R}^{l\times p}$$

是一个深度神经网络, 它的性能 (如泛化性) 自然应是深度 k、宽度 l、每层神经元的非线性传输函数 $\mathcal{G}_i$ 等结构参数的函数. 然而, 如何定量描述或定性地刻画这一函数关系 (即构 – 效关系) 呢? 写出这样的函数可能是艰难的, 但估计它们的性能与结构之间的某种 "可控性" 是可能的. 例如, 让 $\mathcal{E}$ 是一个泛化性度量, $\mathcal{N}^*$ 是理想结构, 则形如

$$L(k, l, G) \leqslant \mathcal{E}(\mathcal{N}) - \mathcal{E}(\mathcal{N}^*) \leqslant B(k, l, G)$$

的不等式估计了深度网络 $\mathcal{N}$ 泛化性能的上、下界. 近年来已有关于深度学习泛化上界的研究, 但还只限于对其中某些单一参数 (如深度) 的影响估计. 更加全面的评估, 特别是有关深度学习泛化下界、本质界的估计尚未见到. 所有这样的研究十分基本, 它不仅能帮助人们认识深度学习机理、评价其性能、改进其结构, 更是设计深度学习网络的理论依据, 是推动深度学习应用从 "艺术" 走向 "科学" 的重要步骤.

建立有确定数学意义的信息深度表示理论是另一个基本数学原理问题. 深度学习的深 "层" 结构代表着它是从深度 "复合" 的意义上来对函数作逼近的, 这使得对深度学习的解释性变得困难. 回想, 数学上的泰勒 (Taylor) 级数展开、傅里叶 (Fourier) 级数展开等, 都为我们提供了非常清晰可解释的函数逼近方式 (如前者以 "逼近阶" 提高的方式, 后者以 "频率" 提高的方式渐近于被逼近函数), 而这些展开是 "叠加" 式的. 所以, 要解释深度网络表示机理, 搭建函数的 "叠加" 式逼近与 "复合" 式逼近之间的桥梁是重要的. 假定 f_k 是对函数 f 的第 k 次近似, ε_k 是对 $f - f_k$ 的某种误差度量, 则

$$f \simeq f_{k+1} = f_k + \varepsilon_k = f_0 + \varepsilon_0 + \varepsilon_1 + \cdots + \varepsilon_k$$

提供了对 f 的一个 "叠加" 式逼近. 根据神经网络的万有逼近定理, 存在线性函数 $L_k(x)$ 和非线性函数 $N_k(x) = G(Wx + b)$ 使它们的复合任意逼近 $f_k + \varepsilon_k$, 即

$$L_k N_k \simeq f_k + \varepsilon_k,$$

$$L_{k+1}N_{k+1} \simeq L_k N_k + \varepsilon_{k+1}.$$

如果我们期望 $L_{k+1}N_{k+1}$ 具有 "复合" 的性质, 即 $L_{k+1}N_{k+1}L_kN_k = L_kN_k + \varepsilon_{k+1}$, 从而它有望通过调整 L_{k+1} 和 N_{k+1} 实现

$$(L_{k+1}N_{k+1} - I)L_kN_k = \varepsilon_{k+1}.$$

这一等式提供了审视深度学习的一个新视角. 一方面, 当深度网络已被训练, 我们可以将上式 ε_k 作为定义, 而通过对 ε_k 的分析来阐明每一层 (块) 的作用 (例如是否单调下降); 另一方面, 可以将 ε_k 设置为优化目标, 对应每一层期望能抽取到的特征 (及由此带来的表示精度变化), 而通过分层目标指导网络训练. 如此能带来一个有确定数学意义、并能明确解释的深度学习架构吗? 这一架构与近年来兴起的残差网 (residual net) 有关联性, 但又明显不同.

学习过程的收敛性也是深度学习一个亟待解决的问题. 深度结构的复杂性 (尤其是各种神经元的非线性) 使得训练一个深度网络是一个高度非线性、非凸的优化问题, 而大数据训练集又使得优化算法的选择离不开随机梯度的使用, 所有这些都使得证明深度学习的收敛性并不容易. 近年来出现了一些通过连续动力系统方法证明深度学习收敛性的尝试, 但很显然, 深度学习训练算法是离散动力系统, 运用连续动力系统方法只能刻画学习率渐近于零时训练算法的收敛性, 实际的深度学习训练算法收敛性还远远没有解决.

深度学习的稳健性也值得深入研究, 它用于揭示当训练集有小的变化时, 网络学习结果是否也会有小的变化. 这一研究对于认识和防止深度学习被欺骗、被攻击有重大意义.

(5) 非正规约束下的最优输运问题

人工智能中的诸多问题都是以数据输运 (data transportation) 或者说数据打通为基础的. 例如, 机器翻译需要把两种语言打通、把语音与文字打通, 机器视觉需要把图像与文字打通, 辅助残疾机器人需要把脑电信号与视觉场景信息打通等. 事实上, 人的认知能力是靠看、听、闻、触等多种感知方式所获得的 "数据" 融合实现的, 这其中所表现的也正是 "把异构的多类数据/信息在某个层面上打通" 这种智能.

数据输运可以形式化为这样的问题: 假定有一种结构的数据 μ_0 和另一种结构的数据集 μ_1, 我们需要在某个约束下将 μ_0 "搬运" 到 μ_1. 让我们用 $\mathcal{F}(\mu_0,\mu_1;\mathcal{P})$ 表示将 μ_0 "搬运" 到 μ_1 且满足约束 $\mathcal{P}$ 的所有可能方式, 则在数学上, 可视 μ_0 和 μ_1 为两个测度, $\mathcal{P}$ 为约束, 可用变换 $T:\mu_0\to\mu_1$ 来实现 "搬运". 于是, 数据输运可建模为如下最优传输问题 (optimal transportation problem, OTP): 寻找 T^* 使满足

$$T^*=\underset{T\in\mathcal{F}(\mu_0,\ \mu_1;\mathcal{P})}{\operatorname{argmin}}\int C\left(x,T\left(x\right)\right)\mathrm{d}\mu_0,$$

其中, $C\left(x,T\left(x\right)\right)$ 表示将 x "搬运" 到 $T\left(x\right)$ 所付出的代价. 当 $C\left(x,T\left(x\right)\right)$ 取为欧氏度量, 即 x 与 $T\left(x\right)$ 之间的欧氏距离, 且 $\mathcal{P}$ 取为 "保质量" 时, 数学上对此已有广泛而深入的研究 (参见文献 [7]), 且已形成一套完整的理论和一些有效的实现算法.

然而, 人工智能的很多应用要求 $\mathcal{P}$ 不是 "保质量", 而是要保其他性质. 例如, 机器翻译要保语义, 医学 CT 转换成 MRI 要保解剖结构, 信息传输从一个网络进入另一个网络要保信息熵等. 对这些非常规约束下的最优传输问题, 无论是数学理论, 还是求解方法都还没有得到研究. 这是人工智能的核心基础问题之一. 很显然, 数据之间之所以需要 "打通", 或者能够 "打通", 根本原因是它们之间存在某些 "共有特征" 或者 "不变量", 如语言翻译之间的语义, CT 转换成 MRI 之间的 "同一人体" 等. "保不变量" 应是数据输运的最本质约束, 含不变量的特征空间是数据输运的可靠 "中间站". 然而, 什么才是一个问题的不变量呢? 一个不变量 (例如语义) 在不同结构空间中 (例如中文语言空间、英文语言空间) 又是如何被表达的? 所给出的两个数据集 μ_0 和 μ_1 各自含有的特征与不变量交集有多大? 如何能够实现 "保不变量" 意义下的最优传输? 所有这些是数据转换、打通的基础, 也是迁移学习的最根本问题.

(6) 如何学习学习方法论

学习方法论是指导、管理学习者如何学习/完成学习任务的一般原则与方法学. 在人工智能从人工化, 走向自动化, 迈向自主化的大趋势下, 让机器学会人类的学习方法论, 或者更严格地说, 学会模拟学习方法论 (simulate learning methodology, SLM) 成为 AI 发展的必由之路 (参见文献 [8]). 作者认为, 学习方法论的模拟可以在不同层次上实现, 例如可通过学习解决一族强相关问题的公共方法论解决另一个强相关问题, 通过学习解决一族强相关问题的公共方法论解决另一个弱相关问题, 通过学习

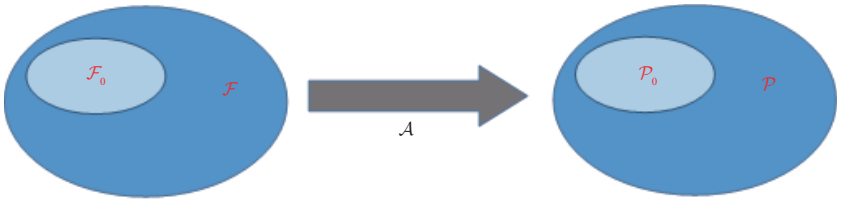

图 1 (网络版彩图) 函数空间上的学习理论

Figure 1 (Color online) A conceptual illustration for learning in infinite functional spaces

解决一族弱相关问题的公共方法论解决另一个不相关问题等. 目前已开始有在第 1 层次和第 2 层次上的探索 (如 learning to learn, learning to teach 等 (参见文献 [9,10])), 但还都集中在非常低的层次上. 目前, 特别需要将学习方法论的学习提升到理论层次. 推动这一提升的关键一步是将学习方法论的学习置入一个合适的数学框架. 假定要解决的问题属于问题类 $\mathcal{F}$, 希望达到的性能是 $\mathcal{P}$. 我们希望通过学习解决 $\mathcal{F}$ 中一个子类 $\mathcal{F}_0$ 问题的公共方法论解决 $\mathcal{F}$ 中的任何一个问题. 假定 $\mathcal{A}$ 是以这种方式解决问题的一个算法. 这种描述提供了学习方法论学习的一个形式化描述, 但还远未构成数学框架, 亟需回答下述理论和实践问题 (参见图 1).

- 如何设计融入学习方法论的问题求解算法 $\mathcal{A}$? (算法构造问题)
- 如何选择供学习的强/弱相关问题集 $\mathcal{F}_0$? (训练集问题)
- 如何刻画从 $\mathcal{F}_0$ 学到的方法论可用于解决更大范围 $\mathcal{F}$ 中的问题? (泛化性问题)

一个显然而基本的问题是: 问题、问题集如何描述呢? 我们认为, 一个问题可以描述为无限维空间中的一个函数 $P(x)$. 换言之, 它是由无限多可能的参数来决定的陈述. 例如, 优化问题 $\mathcal{F}$ 是由凸性、光滑性、非线性等无穷多个性质不同的优化问题所组成的. 深度网络训练问题 $\mathcal{F}$ 由训练集、网络拓扑、损失度量等特征不同的网络训练问题构成. 这样一来, 上述问题就成为函数空间中的学习问题了. 传统学习理论考察对象为数据、训练集和对数据的泛化, 对应地, 函数空间上的学习理论研究问题、训练问题集和对问题的泛化.

函数空间的学习理论是一个尚未得到开垦的领域.

(7) 如何突破机器学习的先验假设

机器学习是人工智能的基础技术, 它所聚焦解决的问题是如何根据输入 – 输出空间 $X \times Y$ 中的数据 $\mathcal{D} = \{x_i, y_i\}$ 对未知输入 – 输出关系 $f : X \to Y$ 作出估计. 机器学习研究的传统范式是: 在一个给定的最优性准则 (即一个损失度量 $l : X \times Y \to \mathbb{R}^+$) 下, 从某个备选函数族 (假设空间) $\mathfrak{F} = \{f_\theta : X \to Y;\ \theta \in \mathcal{P}\}$ 中选择一个函数 f^* 使得它 “平均地” 最优近似 f, 即

$$f^* = \text{argmin}_{f_\theta \in \mathfrak{F}} \{\mathbb{E}_{\mathcal{D}} [l(f_\theta(x), y)] + \lambda p(f_\theta)\},$$

其中 $\mathbb{E}_{\mathcal{D}}$ 表示关于 $\mathcal{D}$ 的数学期望 (均值), $\mathcal{P}$ 是参数集, $p(\cdot)$ 是正则化函数. 机器学习在这样的范式下得到蓬勃发展并构筑了当代人工智能技术的核心算法 (如深度学习、支撑向量机等).

然而, 机器学习的应用与有效性一直是以一些基本的先验假设为前提的 [11]. 例如 (i) 大容量假设: 我们总是假设空间 $\mathfrak{F}$ 的容量是充分大的, 它既能包含 f 但又不依赖于 f 去构造; (ii) 独立性假设: 最优性准则是独立于数据生成机制设置的, 特别是, 损失度量 l 独立于数据而预设; (iii) 完备性假设: 数据 $\mathcal{D} = \{x_i, y_i\}$ 中的标号 y_i 应该是准确无误的, $\{x_i\}$ 是足够多和有代表性的; (iv) 正则子假设: 正则项 $p(f_\theta)$ 的确定是由问题的先验知识决定的, 因而得靠先验事先设置; (v) 欧氏性假设: 机器学习行为是欧氏的, 换言之, 数据集 $\mathcal{D}$ 和参数集 $\mathcal{P}$ 都能够被嵌入到欧氏空间中, 而且学习问题能够在欧氏空间

中分析.

所有这些假设是机器学习研究的常规性设置. 这既可以认为是机器学习能得以飞速发展的原由, 也是机器学习发展的桎梏. 要提高现有人工智能技术的应用水平与性能, 就必须要突破这些机器学习先验假设. 然而, 很容易观察到, 要去掉这些先验假设, 或最优地设置这些参量当且仅当事先知道问题的最优解, 因而问题本身是一个 "先有鸡还是先有蛋" 的问题. 所以, 突破机器先验假设只能是以某种渐近成近似的方式实现的.

大容量假设本质上涉及对待求解问题解空间的刻画. 作者团队所提出的模型驱动的深度学习正是从模型刻画解的角度来解决这一问题的 [12, 13]. 独立性假设涉及如何根据数据集分布特性, 自适性设置学习目标, Meng 等 [14, 15] 所提出的的误差建模原理是突破这一假设的有效方法之一. 完备性假设包括非常复杂的数据集情形, 如非常低质、严重错误、遗失的标签情形, 甚至本身就是小样本. 课程学习、自步学习在解决这一问题上提供了思路 [16, 17], 但如何降低复杂性是必须解决的问题. 突破正则子假设的关键是如何把知识层次的先验通过 "下沉" 到数据层次的学习来实现. 欧氏空间假设涉及选择什么样的数学框架 (例如用什么样的范数) 去分析一个给定的机器学习算法问题. 该问题本质上涉及 Banach 空间几何学与数学上的代数结构 (参见文献 [18]).

(8) 机器学习自动化问题

人工智能的新一轮浪潮在哪? 按照文献 [19, 20] 的判断, 这一新的浪潮应是 "适应环境" 的浪潮. 这一新的浪潮应是在克服现有深度学习只适用封闭静态环境、固定任务、鲁棒性不好、解释性不强等缺陷基础上, 着力发展对开放动态环境可用、稳健、可解释的、自适应的 AI 技术. 美国国防部高级研究计划局 (Defense Advanced Research Projects Agency, DARPA) 2017 年启动的终生学习机项目正反映了这种趋势. 终生学习机涉及两个关键技术领域: 终身学习系统和终身学习自然原则, 前者要求系统可以持续从过程经验中学习, 可以将所学知识应用于新情况, 可以不断扩展自身的能力并提高可靠性; 而后者期望关注生物智能的学习机制, 重点关注自然界生物如何学习并获得自适应能力, 研究上述生物学习原理及技术能否用于机器系统并实际应用. 这一项目大致反映了 "适应环境" 浪潮人们的主要关注点, 也应是人工智能的下一个突破口.

本文认为, 要实现上述这样适应环境的自主人工智能, 一个更为现实而且也是必须实现的中间阶段目标是实现机器学习的自动化 [8]. 机器学习自动化应该解决当今机器学习/深度学习的 "人工化" 和 "难用于开放动态环境" 上存在的突出问题: 数据/样本层面对人工标注的强依赖和对训练样本人工挑选的强依赖, 模型/算法层面对网络结构的预设定和对训练算法的预设定, 任务/环境层面对任务的专属性和对环境的封闭性等. 这样的目标有别于各 AI 公司推出的旨在方便用户挑选模型和参数的 AutoML, 这里的机器学习自动化可称为 Auto^{6}ML, 希望达到如下 "6 个自" 的目标:

- 数据/样本层面: 数据自生成、数据自选择;
- 模型/算法层面: 模型自构建、算法自设计;
- 任务/环境层面: 任务自切换、环境自适应.

实现机器学习自动化与学习方法论的学习有着紧密关联 [8]. 事实上, 我们可以将一个学习任务 $\mathcal{T}$ 理解为一个统计推断, 定义学习空间 $\mathcal{K} = \mathcal{F} \times \mathcal{H} \times \mathcal{L} \times \mathcal{O}$ 是由 4 个基本空间组成的无限维乘积, 其中 $\mathcal{F}$ 是描述数据集的分布函数空间, $\mathcal{H}$ 是机器学习的假设空间, $\mathcal{L}$ 是损失函数空间, $\mathcal{O}$ 是优化算法空间, 从而, 一个机器学习方法可被定义为 $\mathcal{K}$ 上的一个赋值 (用超参数化的语言, 该赋值为 4 个无穷序列值); 进而我们可定义学习方法论是一个从任务空间 $\mathcal{T}$ 到学习空间 $\mathcal{K}$ 的映射 $\mathcal{LM} : \mathcal{T} \to \mathcal{K}$, $\mathcal{LM}(\mathcal{T})$ 称为一个超参赋值. 这样, 学习方法论学习 (SLM) 便可理解为是一个对超参规则的一个学习问题. 一旦这样的规则被学习到, 通过固定 $\mathcal{F} \times \mathcal{H} \times \mathcal{L} \times \mathcal{O}$ 中的部分要素而优化其他要素, 便能实现 Auto^{6}ML

中的某一或某些自动化目标. 所以, 学习方法论学习可能为机器学习自动化的实现提供模型和算法基础.

(9) 知识推理与数据学习的融合

人工智能研究已经经历了 “手工知识” 和 “统计学习” 两次浪潮, 现正进入 “适应环境” 的浪潮. 第 1 次浪潮以符号推理/知识库运用为特征, 知识表示需要人工设定, 对少数特定领域的知识推理能力强, 但感知能力弱; 第 2 次浪潮以基于数据/机器学习为特征, ANN 广泛使用, 知识自动表示, 对特定领域感知和学习能力强, 但抽象和推理能力差; 第 3 次浪潮会以自主学习/适应环境为特征, 不会仅仅是前两次浪潮能力的简单叠加, 将会具备持续自主学习能力, 抽象能力也会大幅提升.

在这样的发展大趋势下, 要求所研发的 AI 系统既具有强大的知识自表示/自学习功能, 又具有强大的知识推理功能是自然不过的事. 特别是, 后深度学习时代必然追求把知识推理与深度学习能够结合起来, 以使得深度学习在保持强大的数据学习能力基础上, 具有更明确的可解释性和更强的泛化性. 在这一努力中, 一个融合数据学习和符号推理于一体的框架或者模型是基础. 这样的模型应该能够同时处理两类不同的变量 — 数据变量 (连续或离散的实数形式) 和逻辑变量 (符号形式), 以及处理两类不同的运算 — 实数运算和逻辑运算. 设计和运行这样的融合系统核心困难在如何联通数据蕴含的知识和语言表达的知识, 其中知识的数据化 (数字化) (从抽象到具体/示例) 和数据的知识化 (从具体/示例到抽象) 都是必须解决的问题. 在数据空间和语义空间之间建立一个中间空间来联通数据与知识应该是一个很好的选择 (参见文献 [21]).

理论上分析数据 – 知识混合系统也是一个挑战性问题.

(10) 智能寻优与 AI 芯片

寻优是最典型的智能行为之一, 也是 AI 的最重要应用场景之一. 设计一个好的深度网络架构、选择一个合适的训练算法 (如确定 BP 算法的学习率)、挑选合适的训练数据、保证学习过程的收敛性/泛化性等, 这些与深度学习应用相关的问题自身就是复杂的寻优问题. 科学、工程、管理领域所广泛出现的各种复杂约束下的组合设计, 不确定环境下的对抗博弈, 难以解析表达的复杂优化等也都是典型的寻优问题.

数学的最优化理论与方法提供了寻优的理论基础, 但数学方法通常假定目标函数的解析形式、具有凸结构和超参数已知. 解决寻优问题也并不是深度学习的特长. 所以, 如何设计寻优的人工智能过程仍远未解决. 正如 AlphaGo [22] 所展现的, 设计高效的这样一种寻优算法会是一个高度综合的系统, 可能是随机搜索与确定性规则导航的结合, 是数学优化算法与深度学习方法的结合.

已经出现各种各样模拟具有内在收敛特性和进化特征的自然算法, 如群体智能算法 (遗传算法、蚁群算法、粒子群优化、烟花算法等), 模拟退火算法, 文化算法等. 这些算法有很深刻的自然/生物/社会/物理解释, 有天然的并行性和对目标函数的高度容错性 (可以不知道目标函数的解析形式), 但普遍缺少严密的理论分析, 且收敛极其缓慢. 如何将这类算法与深度学习结合, 形成具有学习能力而且真正高效的新一代自然算法值得期待.

模拟生物/物理等自然过程的寻优策略也最好以生物/物理的方式实现. 这既是一个朴素的但同时又是未来可能突破的方向. 例如, 模拟退火算法是模拟金属通过退火工艺实现最优的晶体结构的算法, 已经知道, 这是一个理论上保证收敛的全局优化算法. 退火是一个加热固体然后缓慢冷却至结晶的过程, 其本质是改变系统所处的温度场, 所以, 是能够物理模拟的 (例如通过改变光强、磁场强度等). 受量子力学启发, 基于量子比特机制的量子退火方法已经出现, 并已被成功研制成数字退火芯片. 这种数字退火芯片在 CMOS 数字电路上再现量子比特机制, 退火速度极快, 不再需要像传统计算机那样编程, 只需通过简单设置参数来执行计算.

组合优化问题求解一直是寻优的难中之难. 推广 AlphaGo 算法到一般组合优化问题 (如 TSP) 原理上是可能的, 但如何克服问题维度的可变性和训练数据生成是必须克服的困难. 在传统计算模式下解决组合优化问题遭遇根本困难, 将数字退火与描述一般组合优化问题的伊辛 (Ising) 模型结合可能带来突破. 2007 年, 加拿大 D-wave 首创了量子退火机 (参见文献 [23]), 该机器将量子退火用于伊辛模型并通过 “退火” 场环境下使用叠加态搜索各种可能性解决组合优化问题. 量子退火机将伊辛模型中的节点处理为量子态 (相当于一个量子比特), 整个计算可通过非量子范式计算, 因而有广泛的应用前景. 深度学习依赖超快速和大规模的矩阵乘法及累加, 这种操作可以在量子退火机上执行. 量子启发算法及其专用芯片, 为解决当前基于冯 · 诺依曼架构的计算机很难解决的实际组合优化问题带来了曙光, 也必将进一步推动技术的发展, 并在广泛的应用领域带来重大突破.

人工智能芯片是加速 AI 算法执行的利器, 其功能是将 AI 算法中特定重复使用的运算/操作硬件化 (以片上集成电路的方式), 本质是以物理的方式实现数学运算或操作. 这种加速数学算法的物理措施不仅是提高人工智能应用效率而且也是突破计算科学中一些瓶颈问题的关键. 基于矩阵乘法及累加的人工智能芯片已经出现, 但它的存算一体实现, 尤其是高可靠、高精度、低功耗仍需要进一步突破. 人工智能芯片从专用走向通用是一种必然, 这其中的关键是合理抽象人工智能算法中最为普遍使用的数学和逻辑操作, 并予以物理化实现. 基于某些更高层次的优化算法 (例如 ADMM) 芯片来加速人工智能算法是值得探讨的方向.

3 结论

上述 10 个人工智能重大数理基础问题已在国内外引起高度关注, 例如, 问题 (1) 和 (10) 已作为国家自然科学基金委重大项目立项研究, 问题 (2), (4), (5), (8) 已作为科技部变革性技术关键科学问题和数学与数学应用重大专项立项研究. 对于这些问题的研究, 国内外不同领域也已经正在取得重要进展. 例如, 统计学界近年来有关高维、稀疏、分布式统计方面的研究取得了突破性进展; 数学界有关深度学习泛化性、深度学习与微分方程数值解的关联性等方面取得了重要进展; 机器学习界在突破机器学习先验假设、开拓新的学习范式上取得了持续的重要进展; 人工智能领军企业在突破应用系统和研发 AI 芯片方面也取得了重大进展. 但是, 必须注意到, 所有这些进展都还远远没有解决所提出的 AI 数理基础问题. 解决这些重大的数理基础问题构成了人工智能未来发展的驱动力和重要前沿领域.

参考文献

1 The National Research Council of the United States. Revitalizing American mathematics — a plan in the 1990s. Beijing: World Book Publishing Company Beijing Branch, 1993 [美国国家研究委员会著. 叶其孝, 刘燕, 章学诚, 等译. 振兴美国数学 —— 90 年代的计划. 北京: 世界图书出版公司北京分公司, 1993]

2 Xu Z B. To make good use of big data requires great wisdom — accurately grasp and scientifically respond to the opportunities and challenges brought by big data. People’s Daily, 2016-03-15 (07) [徐宗本. 用好大数据须有大智慧 —— 准确把握、科学应对大数据带来的机遇和挑战. 人民日报, 2016-03-15 (07)]

3 Xu Z B. Grasp the focus of the new generation of information technology: digitalization, networking, and intelligence. People’s Daily, 2019-03-01 (09) [徐宗本. 把握新一代信息技术的聚焦点: 数字化、网络化、智能化. 人民日报, 2019-03-01 (09)]

4 Xu Z B, Tang N S, Cheng X Q. Data Science—Its Connotation, Method, Significance and Development. Beijing: Science Publishing. 2021 [徐宗本, 唐年胜, 程学旗. 数据科学 —— 它的内涵、方法、意义与发展. 北京: 科学出版社. 2021]

5 Chen H H, Wang D H. Fan Jianqing: mathematics as a tool to solve social problems. Science Times, 2006-12-14 [陈欢欢, 王丹红. 范剑青: 把数学作为解决社会问题的工具. 科学时报, 2006-12-14]

6 National Research Council. Frontiers in Massive Data Analysis. Washington: National Academies Press. 2013
7 Villanic C. Optimal Transport: Old and New. Berlin: Springer Science & Business Media, 2008
8 Shu J, Meng D Y, Xu Z B. Learning an explicit hyperparameter prediction policy conditioned on tasks. 2021. arXiv:2107.02378v1
9 Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, 2016. 3988–3996
10 Arends R I. Learning to Teach. 5th ed. New York: The McGraw-Hill Companies. 2000
11 Xu Z B. How to break through the prior assumptions of machine learning? In: Proceedings of International Artificial Intelligence Technology Conference Speech Report, 2021 [徐宗本. 如何突破机器学习的先验假设? 见: 2021 全球人工智能技术大会演讲报告, 2021]
12 Xu Z B, Sun J. Model-driven deep-learning. Natl Sci Rev, 2018, 5: 22–24
13 Yang Y, Sun J, Li H B, et al. ADMM-CSNet: a deep learning approach for image compressive sensing. IEEE Trans Pattern Anal Mach Intell, 2020, 42: 521–538
14 Meng D Y, Torre F D L. Robust matrix factorization with unknown noise. In: Proceedings of IEEE International Conference on Computer Vision, 2013
15 Zhao Q, Meng D Y, Xu Z B, et al. Robust principle componenent analysis with complex noise. In: Proceedings of the 31st International Conference on Machine Learning, 2014. 55–63
16 Ma F, Meng D Y, Dong X Y, et al. Self-paced multi-view co-training. J Mach Learn Res, 2020, 21: 1–38
17 Wang K D, Wang Y, Zhao Q, et al. SPLBoost: an improved robust boosting algorithm based on self-paced learning. IEEE Trans Cybern, 2021, 51: 1556–1570
18 Xu Z B, Roach G F. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. J Math Anal Appl, 1991, 157: 189–210
19 Tan T N. The history, current situation and future of artificial intelligence. Qiu Shi, 2019, 4: 39–46 [谭铁牛. 人工智能的历史、现状和未来. 求是, 2019, 4: 39–46]
20 Yang X J. A brief history of intelligence. In: Proceedings of Artificial Intelligence Frontier Technology Forum, Changsha, 2017 [杨学军. 智能简史. 见: 人工智能前沿技术论坛, 长沙, 2017]
21 Zhang B, Zhu J, Su H. Toward the third generation of artificial intelligence. Sci Sin Inform, 2020, 50: 1281–1302 [张钹, 朱军, 苏航. 迈向第三代人工智能. 中国科学: 信息科学, 2020, 50: 1281–1302]
22 Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. Nature, 2017, 550: 354–359
23 Zhang C X. AI CHIPS—Cutting-Edge Technologies and lnnovative Future. Beijing: Posts & Telecom Press, 2021 [张臣雄. AI 芯片 —— 前沿技术与创新未来. 北京: 人民邮电出版社, 2021]

Ten fundamental problems for artificial intelligence: mathematical and physical aspects

Zongben XU[1,2,3]

1. *Xi'an Academy of Mathematics & Mathematical Technology, Xi'an Jiaotong University, Xi'an* 710049*, China*;
2. *Pazhou Lab, Guangzhou* 510335*, China*;
3. *Peng Cheng Laboratory, Shenzhen* 518055*, China*
E-mail: zbxu@mail.xjtu.edu.cn

Abstract Ten fundamental to-be-solved problems for current artificial intelligence research and development are proposed and analyzed from the mathematical and physical points of view. These 10 problems include: (1) the statistical foundation problem for big data analysis, (2) the algorithm problem for basic mathematical computation in big data circumstances, (3) the understanding of structures and properties of various data spaces, (4) the mathematical theories on deep learning, (5) the investigation on the optimal transportation problem under irregular constraints, (6) the problem of how to simulate learning methodology and learn in infinite functional spaces, (7) breaking through the prior hypotheses of machine learning, (8) realizing machine learning automation from selection to creation, (9) integrating knowledge inference and data learning, (10) intelligent search and AI chips.

Keywords artificial intelligence, mathematical and physical foundation of AI, statistics, big data algorithms, data spaces, deep learning, optimal transportation problem, simulate learning methodology, hypotheses on machine learning, machine learning automation, AI chips

Zongben XU received his Ph.D. degree in mathematics in 1987 from Xi'an Jiaotong University, China. He is currently the director of National Lab for Big Data Analytics, the dean of Xi'an Academy of Mathematics and Mathematical Technologies, and the director of PaZhou Lab, Guangzhou. His research interests mainly include intelligent information processing, machine learning and theories in numerical modeling.