艾伦·图灵著《计算机器与智能》的要点归纳与启发

刘宇航[1] 孟轩宇[2]
[1] 中国科学院计算技术研究所
[2] 深圳外国语学校

关键词：图灵 智能

概述

人类的历史现在走到了2020年，人工智能之前经历了多次兴衰起伏，当前正处于一个蓬勃发展的阶段（主要归功于计算能力和存储能力的提高）。回顾经典文献，可能有助于找到现阶段的定位，回归基本问题，追根溯源，看看已经取得了哪些进展以及还有哪些差距，思考未来可能走多远以及可能的前进方向。

《计算机器与智能》(Computing Machinery and Intelligence) 是艾伦·图灵 (Alan Turing) 于整整70年前，即1950年在《心智》(*Mind*) 杂志上发表的在计算机科学发展史上具有奠基意义的一篇经典文章。《心智》是心理学和哲学领域的期刊，每个季度出版一期。图灵这篇论文是1950年第4季度这期的第1篇文章。这篇文章原文共28页，分为7章：(1) 模仿游戏，(2) 对新问题的评价，(3) 游戏中的机器，(4) 数字计算机，(5) 数字计算机的通用性，(6) 主要问题的对立观点，(7) 学习机器。全文按照“提出问题—分析问题—解决问题”的顺序展开，在分析和解决问题时，先在第6章反驳9种对立的观点，然后在第7章提出正面的观点，讲述实现的思路。

图灵出生于1912年[1]，在《计算机器与智能》一文发表时只有38岁。在此之前，图灵于1936年发表了《论可计算数及其在判定问题上的应用》(On Computable Numbers, with an Application to the Entscheidungsproblem)，那时他只有24岁。一个24岁的年轻人，在现代计算机还有10年才诞生的时候，去想象一个机器（被后人称为“图灵机”），论证可计算性，提出“有些问题不可以被机器计算”；一个38岁的年轻人，在计算机刚刚起步（只有4年）的时刻，讨论“计算机器与智能”（被后人称为“图灵测试”），提出“有可能构建具有智能的机器”，显示了弥足珍贵的创新能力，人们应该鼓励、赞赏，而不是讥笑、反对。

图灵文章的标题包含两个非常基本的名词，一个是计算机器 (computing machinery)，一个是智能 (intelligence)，即使在70年后讨论这两个基本概念仍很容易陷入空谈，或者原地踏步、表面很热闹但却是循环论证，或者“盲人摸象”，总之不容易取得实质性的、全面的、正确的新结果。什么是计算机？什么是智能？计算机能否思考？这样的“基本”问题值得回答吗？能够回答吗？

有的人不仅仅对问题的答案有不同意见，对问题本身也有不同意见，他们认为：不要“正名”，只管用就好了；或者说，只去做具体的设计和应用就

[1] 巧合的是，图灵与我国的“三钱”（钱学森、钱伟长、钱三强）几乎同龄。

好了，不要思考“空洞”“好高骛远”“虚无缥缈”的问题；或者说，只关注怎样做就好了，不要关注为什么和是什么。持这些观点的人看到图灵撰写的《计算机器与智能》或许应该有所反思。

要点一：用多学科的不同角度研究智能

图灵具有扎实的物理学基础，他提到了经典物理学中拉普拉斯 (Laplace) 的确定论观点。图灵熟知计算机发展历史，对巴贝奇 (Babbage) 分析机和曼彻斯特机具有深入的了解。

全文内容丰富，涉及神学、物理、化学、生物、信息论、数理逻辑等多个学科，但紧扣“机器是否能够思考”这一主题，可以看出图灵能够得心应手地驾驭多学科的知识。人类生活在一个大数据、“知识爆炸”的时代，但如何驾驭知识却成了难题。要鼓励人类培养独立思考、跨学科思考的精神，改进和革新教育形式和内容，这在人工智能时代具有重要意义。

图灵首先批驳的是神学的观点：“思维是人类不朽灵魂的一项功能。上帝只赋予每个善男信女不朽的灵魂，但从未将之赐予任何其他的动物或机器。所以，动物或者机器不能思维”。图灵采取科学的态度，中立、谦卑，但不失锐利，有判断但不武断，能用神学的语言去换位思考，回复神学的观点。神学认为只有人才被赋予灵魂，动物或机器均没有被赋予灵魂。图灵首先从距离的角度去考虑问题[2]，D（动物，人）<D（生物，非生物），一下子提出了神学观点的改进版。当一个观点被改进之后更合理，则说明这个观点被改进之前具有不合理性。实际上，我们可以展开去思考。D（动物，人）<D（动物，植物）<D（植物，非生物）。我们可以推断，D 函数肯定是与智能相关的一个重要函数。其次，神学本身不是铁板一块，图灵指出“伊斯兰教认为妇女没有灵魂，基督教对此有何感想？”，显然神学不同派别之间存在教义上的矛盾。最后，图灵指出，既然认为上帝是万能的，那么上帝为什么不能赋予动物灵魂呢？

第二个批驳的是“鸵鸟”式的观点：“机器能够思考将导致严重的后果。让我们希望和相信机器不能思考”。图灵认为这一观点不够实在 (substantial)，所以无须一驳。科学的结论或结果是不以人的意志为转移的，“鸵鸟”式的观点显然不是科学的观点。值得指出的是，对“鸵鸟”式观点的异议在人工智能发展较为顺利的时期（比如在当前）往往比较热烈，主要是担心人类的优先性和优越感被挑战。现在美国人所倡导的“美国优先”也是这种思维，他们不愿意看到具有强大国力的中国，就像“鸵鸟”式的人不愿意看到具有强大智能的机器一样。

第三个批驳的观点来自数学：根据哥德尔定理以及丘奇、克莱恩、罗瑟、图灵的不可判定定理，任何离散状态机的能力都是有限的，所以机器不能够思考。图灵的反驳是：任意一台特定机器的能力都是有限的，人的能力也可能是有限的；机器有时会犯错误，但这没什么，因为人也经常犯错误。人类犯了太多的错误，所以没有资格因为机器犯错误而产生优越感。

诸如此类，图灵还批驳了其他 6 种观点[3]。发展智能要秉持科学的态度，而不是盲目乐观或悲观。图灵具有非凡的思考力，能列举 9 个方面的对立观点，然后简明扼要地进行有力的批驳。那么能否列举第 10、11 乃至更多的对立观点呢？读者可以进一步思考。

要点二：定义概念是分析问题的基础

图灵是务实的，他以严谨的态度做科学研究，而不是在创作科幻小说。他在使用概念之前对概念做出了清晰的界定。

他对“机器”的含义做了界定。(1) 在一个现实的机器上，不可能同时使用人类的一切技术或大多数技术。图灵在文章中没有把 DNA 计算、量子计算等技术作为基础，而是将电子计算机或数字计算机作为基础。(2) 能制造机器，不代表能清晰准确地

[2] 这个 D 函数是我们根据图灵原文的意思提出的。

[3] 限于篇幅，我们不在这里逐一介绍和分析这些对立意见，而是选择在一本即将出版的专著中详细论述。

描述机器。通过实验方法制造的机器与采用递归函数等数学方式构思的机器是不同的。(3) 为了讨论的方便，图灵将人排除在机器的定义之外[4]。

图灵在文章第 1 章认为需要对“思考”做出定义。图灵的定义是：如果机器能在模仿游戏中表现出色，机器就可以被认为能够思考，或者说具有智能。图灵把原问题“机器能思考吗”转化为一个新问题“机器能在模仿游戏中表现出色吗？”，然后在第 6 章第 2 段明确提出原问题是无意义的，不值得讨论。

要点三：通过思维实验研究智能

图灵批判了一种流行的观点：“科学家进行科学研究工作总是从可靠的事实到可靠的事实，从来不受任何未经证明的猜想所影响”。图灵指出“这种看法实际上是相当错误的。假如能清楚哪些是经过证明的事实，哪些是猜想，将不会产生害处。猜想是极其重要的，因为它们能提示有用的研究线索。”

图灵在这里做的是一个关于存在性的思维实验，就像伽利略关于自由落体的思维实验一样[5]，而不是一个现实实验。思维实验是指使用想象力去进行的实验，所做的都是在现实中无法做到（或现实未做到）的实验。证明 (proof) 与证实 (verification) 是有重要区别的。

图灵指出，不要求所有的数字计算机都能在模仿游戏（即图灵测试）中表现良好，即使 99% 的数字计算机表现不合格，只要有一台表现合格就可以了；也不要求现在的数字计算机在游戏中表现良好，只要 10 年甚至 100 年以后的数字计算机在游戏中表现良好即可。这样的思想与图灵 1936 年发表的《论可计算数及其在判定问题上的应用》关于可计算性的思想是一致的：存储空间是无限的，时间是无限的，在这个意义上讨论可计算性。

有不少人认为只要存储空间是无限的，时间是无限的，问题总是能够算完，这是错误的。实际上即使假设存储空间是无限的，时间是无限的，仍有很多问题是不可计算的；何况在现实中存储空间和计算时间有限，不可计算的问题更多。以存储空间和时间均是无限为前提的可计算性，称为图灵可计算性；以存储空间和时间均是有限为前提的可计算性，称为现实可计算性。图灵不可计算的，一定是现实不可计算的；图灵可计算的，未必是现实可计算的。

要点四：通过对人类与机器作类比和对比来研究智能

我们注意到，图灵在标题中说计算机时，没有用“computer”这个词，因为这个词从本意上可以指从事计算的人 (human computer)。虽然今天，computer 一般指计算机，但为了避免歧义，图灵使用“computing machinery”来指代从事计算的机器，是非常精准和恰当的。

图灵在文章中提出了人类计算机 (human computer) 和数字计算机 (digital computer) 的概念。在今天的关于人工智能的讨论中，“人类计算机”这个词不被常用，这是不应该的，人类计算机与数字计算机，人类智能与人工智能，都是相互参照和对应而存在。人类计算机的“规则书”(book of rules) 对

[4] 图灵将人与机器这两个术语分开的做法是睿智的，事实上关于人与机器之间的关系，存在多种不同甚至截然相反的观点。法国科学家拉·梅特里认为“人是机器”，图灵奖得主明斯基认为“大脑不过是肉做的机器而已”(The brain happens to be a meat machine)，另一位图灵奖得主威尔克斯认为“动物和机器使用完全不同的材料，按十分不同的原理构成”(Animal and machine are constructed from entirely different materials and on quite different principles)。

[5] 亚里士多德认为，越重的物体下落越快。伽利略做了这样一个思维实验：假设有一个重量为 8 的物体，另一个重量为 4 的物体。那么，重量为 8 的物体下落应该比重量为 4 的这个物体下落快。如果我们把两个物体用绳子牵在一起，由于速度慢的那个物体对快物体的牵连，二者连在一起下落的速度应该介于二者单独下落时的速度之间。但是，换一个角度考虑，重量分别为 8 和 4 的两个物体连在一起，可以被视为一个重量为 12 的物体，那么，根据亚里士多德的观点，这个新物体的速度应该比刚才的两个物体都快！这与我们刚才得到的结论（速度介于二者之间）是矛盾的。如果去做真实实验，会受到空气阻力、计时设备等各种干扰，真相可能被掩盖或歪曲，伽利略用他巧妙的思维实验解决了这个问题。

应于数字计算机的程序 (program)。

对人来说，时间、耐心是稀缺资源。计算机就是用来弥补人的这些稀缺资源。计算机相对于人，在运算的速度和精度、重复做某件事的耐心上有优势。

图灵介绍了指令的格式、顺序执行和跳转执行，特别详细地介绍了循环结构。机器的一个优势是擅长做重复的操作，弥补人类耐心（是一种稀缺资源）的不足。循环是程序中表达语义的重要程序结构，也往往是最耗时的部分，因此往往成为性能瓶颈，被称为“热点”。

要点五：编程与存储在智能中具有重要作用

从反面和正面两个角度讨论问题，既要善于破坏一个旧世界（反驳对立的观点），又要善于建设一个新世界（给出构建智能的方法）。在原文最后一章，图灵从正面论证“有可能存在可以思考的机器”。

图灵指出了编程 (programming) 的实质：“给一个机器编程使之执行操作 A”，意味着把合适的指令表放入机器以使它能够执行 A，从这个意义上看，是人类通过编程把自己的智能注入 (inject) 了机器。当前“赋能”这个词汇被广泛使用，图灵的这篇文章的“注入”一词应该是“赋能”的本源出处。操作 (operation) 是一个落实执行计算 (computing) 的实体。指令集体系结构是软件与硬件的接口。规则书、指令表都是指程序，控制器是解读和遵循程序的实体，显然程序与智能之间存在很强的关联。图灵强调了编程对智能的重要性。

图灵估计，50 年后（即 2000 年）计算机的存储容量是 1Gb，所以计算机在模仿游戏中的表现会更好。但他并没有预言计算机在 2000 年就可以具有思维、能够思考了。他的表述是“一般提问者在提问 5 分钟后，能准确判断的概率不会超过 70%”。这是一个很睿智的表述。我们把这个表述一般化，“一般提问者在提问 n 秒后，能准确判断的概率不会超过 m%”。n 越大、m 越小，机器的智能水平越高。这里实际上考虑了性能、响应时间的因素，所以含有“实用可计算”（由中科院计算所研究员徐志伟定义[6]）的思想。

无限容量的计算机是不存在的，过去、现在、将来都不存在，但具有特殊的理论价值。而存储容量是可以逐步扩展的。从这里可以看出，不存在的东西不一定没有价值。机器的存储容量大，意味着机器可能拥有的状态数量很大。图灵提到了计算机的三个部分：存储器、执行单元和控制器，没有提到输入设备和输出设备。存储器中存放的是数据和程序，其中数据是程序的处理对象，数据可被分为初始数据、计算过程中产生的中间数据和计算过程结束产生的最终数据。一个人（人类计算机）具有很强的心算能力，通常是指能够在没有纸、笔等工具的条件下进行计算，这就需要这个人具备很强的记忆力，特别是关于计算过程中的中间数据的记忆能力，更具体地说，是对应于高速缓存 (cache) 的那部分短时记忆能力。

要点六：研究智能要抓住本质属性

图灵文章的第 2 章标题是“对新问题的评价”。图灵指出“外表与智能无关”。一个外表很像人的机器，可能只有很低的智能；一个外表不像人的机器，可能具备很高的智能。智能是各种能力中的一种，除了智能，还有力量（爆发力、耐力）、听觉、视觉、味觉、嗅觉、运动（敏捷性、持久性）、魅力、勇敢等各种能力属性。有智能不代表一定全能，某些维度上无能也不代表没有智能或智能低下。

图灵注重把握本质，他说“我们不希望因机器不能在选美比赛中取胜而被惩罚，正如我们不希望因人不能在和飞机赛跑中取胜而被惩罚一样，我们的游戏设定让这些无能变得无关紧要”。机器是否用电，对本质没有影响。通过化学过程、电、机械，都可能实现等价的数字计算机。图灵再次强调对非本质的属性（腿和眼）进行忽略。

马克思说“人的本质属性在于人的社会性”。

[6] 徐志伟，李春典 . 低熵云计算系统 . 中国科学：信息科学，2017(09):27-41.

对新生事物如何理解，一种比较常见的做法是组织一些专家分别给出意见，然后按照“少数服从多数”这样的盖洛普统计方式做出决定。图灵指出这种做法是危险的。他把问题等价转化了，这是一个具有非凡创造性的做法。论文第 1 章“模仿游戏”涉及到“虚拟化在智能中具有什么作用？”“功能和性能有何区别？”这样的本质问题。

要点七：智能与泛化能力、随机性、奖励机制有关联

图灵具有扎实的物理学基础，能够得心应手地驾驭原子的裂变反应、临界体积这些知识，表现出极高的人类智能。人类现在生活在一个大数据时代，数据不等于知识，知识不等于智能，“填鸭式”的应试教育将越来越不能适应时代的要求，重要的是培养学生的分析能力、洞察能力、联想和想象能力。

图灵用原子堆来比喻人脑：“如果原子堆的规模变得足够大的时候，轰击进来的中子产生的扰动很可能会持续地增加，直到整个原子堆解体。思维中是否存在一种对应的现象呢？机器中呢？这样的现象在人类头脑 (human mind) 中应该是存在的。绝大多数头脑都处于‘亚临界’(sub-critical) 状态，对应处于亚临界体积的反应堆。一个想法进入这样的头脑中，平均下来只会产生少于一个的想法作为回复 (in reply)。有一小部分思维处于超临界 (super-critical) 状态，进入其中的想法将会产生二级、三级甚至越来越多的想法，最终成为一个完整的‘理论’。动物的头脑看起来肯定是处于亚临界状态的。从这种类比出发，我们要问：‘机器可以被制造成超临界的吗？’”。能不能举一反三、触类旁通，是智能高低的一个判断准则。现在机器学习中的一个词“泛化能力”就是说这个事情。

图灵在原文中多次提到随机数和随机算法，说明智能与随机性、不确定性之间存在着关联。顺便提一下我国的两位人工智能科学家，亚裔唯一 ACM 图灵奖得主姚期智研究的就是伪随机数生成理论，中国工程院院士李德毅研究的是不确定性人工智能。

图灵指出，需要将惩罚和奖励与教学过程联系在一起，一些简单的儿童机器可以按照这种原则来构建或编程，使得遭到惩罚的事件不大可能重复，而受到奖励的事件则会增加重复的可能性。惩罚和奖励是教学过程中所必需的。教学的过程本质上是训练的过程，也就是学习的过程。“使得遭到惩罚的事件不大可能重复，而受到奖励的事件则会增加重复的可能性”，这不就是神经网络反向传播的基本原理吗？现代医学认为，大脑内部存在一个奖励机制，很多药物、毒品成瘾都与这个机制有关。

结束语

图灵在原文最后指出，他讨论的是机器与人在“所有纯智力领域”竞争。需要注意两点，第一，不是在非智力领域竞争，例如人类不如挖掘机那样有很强大的臂力，但这没有讨论的必要。第二，不是仅仅在部分智力领域竞争，而是所有智力领域竞争，比如机器与人不仅仅要比赛下棋，还要在写十四行诗等其他领域比赛。

图灵是一个具有深邃思考力的远望者，但他深知一个人的视距是有限的。图灵原文的最后一句非常有深意，一方面是文章结束语，表示即使不远的将来，也有很多工作要做；另一方面，这是人工智能的思考范式，下棋时，我们只是看到不远的前方，但即使这样，这中间也有很多选择。 ■

刘宇航

CCF 专业会员，*CCCF* 特邀译者、特邀专栏作家。中科院计算所副研究员、硕士生导师。主要研究方向为计算机体系结构、高性能计算、大数据、智能并发系统。
liuyuhang@ict.ac.cn

孟轩宇

深圳外国语学校国际部九年级学生，曾在中国科学院数学与系统科学研究院和计算技术研究所参加学习，在数学建模、单纯形算法、计算机性能分析、人工智能和大数据等领域有过较深入的学习。