图灵1951年报告Intelligent Machinery, A Heretical Theory汉译文

经典人工智能文章《智能机器，被视为异端的理论》（Intelligent machinery, a heretical theory）来自于1951年，阿兰·图灵在BBC电台节目The'51 Society中为大众做的报告[1]，旨在介绍“智能机器（Intelligent machinery）”的原理。

在文章中，图灵用“learn byexperience（通过经验学习）”来阐释“智能机器”：从一个相对而言较为简单的机器来开始这一切，通过让其经历一系列所谓的“经验”，逐步将其转化为一套精巧得多的机器，而且还能够被用来处理更为广泛的各类偶发事件。[2]

图灵这里所说的“智能机器”与人工智能的核心“Machine Learning（机器学习）”密切相关，希望图灵这篇文章能追本溯源为理解“机器学习”这一基本而复杂的概念带来启发性的帮助。

这里刊载的全文由刘德欣主译，部分译文及解读见智能哲学：逆源图灵的思想看人、机关系。

参考文献：

［1］Alan Turing, Intelligentmachinery, aheretical theory，1951，http://viola.informatik.uni-bremen.de/typo/fileadmin/media/lernen/Turing-\_Intelligent\_Machinery.pdf

［2］B. Jack. Copeland, The Essential Turing,2004

智能机器，被视为异端的理论Intelligent machinery, aheretical theory

A.M.图灵，1951年BBC的The'51 Society电台节目

“你不可能制造出一台可以为你思考的机器”，这似乎是一个已经被普遍接受的不再被质疑的观点。

本文的目的就是要质疑这个似乎已经被广为接受的观点。

大多数为商业用途开发的机器是被用于完成一些特定领域的特定工作，并且以可观的速度获得确定的结果。通常情况下，机器被用来完成一系列同样且重复的工作。这个基于我们实际所能得到的机器的这个事实，成为了很多持如上标语所示观点的人们的一个强大论据。然而，对一个数理学家来说，这个论据是明显站不住脚的。因为有证据表明，存在这样的机器，理论上是可以用非常接近人类思考的方式来完成一些事情。举例来说，这种能力可以被用来检验《数学原理》体系当中出现的正式证据的有效性，甚至可被用来判定这个《数学原理》体系当中的公式到底是可被证明为真还是被证明为假。但是，在公式本身不能被证实真假的情况下，这样的机器的表现肯定是不能够令人满意的。因为它会持续不断工作，而根本不产生任何有效结果，但这一点不能被视为与数学家在类似情境中的反应有很大的不同，比如对照在过去数百年间，数学家在证明费尔马大定理之真伪的过程的实际表现。对这类机器来说，一个更为精巧微妙的论据支持无疑是必要的。根据著名的哥德尔定理，或者一些类似的根据，人们可以证明：尽管这类机器可以被建造出来，但肯定有一些事例，机器对此是不可能给出一个答案的，但一个数学家却可以。从另外一个角度来说，机器的确拥有一些相对于数学家的优势。只要不出现所谓的机械“系统崩溃”，那么其实际所为是可以被信赖的，而与此相对照，数学家却总是会犯一定比例的错误。我倾向于相信数学家会犯错的这种所谓的“危险”的一个不可避免的必然产物，正是导致其有时会偶然获得一种全新的解决问题的方法的原因。这一点似乎可以从一个更广为众人所知的事实方面来得到确认，既最可靠的人往往不能获得解决问题的新方法。

我争论的观点是：那种能够高真模仿人类行为的机器是可以被造出来的。它们有时会犯错误，有时会得出十分新颖有趣的结果，总的来说，它们的结果输出可以与人类的输出一样,被予以等量齐观的关注。这个结果的内容在于对真实结果的更高的概率期望，而对于这一点，我认为它恰恰不可能被给予一个精确的描述。例如，你不能因为一台具体的机器迟早可以得出所有可能的结果就简单地说机器迟早会得出任何正确的结果。我们也知道如何去建造这种机器，它们会以基本相同的概率输出正确的或错误的结果，如果是这样，那么它们输出的结果是没有任何意义的。而且如果这真的可以被证明的话，那么机器对客观情境的真实反应会证明我的论点。

让我们更深入细致地进入到“证据”本身。制造一台在任何程度的测试下都能得到良好结果的机器是完全可能的，条件是机器被做得足够精细。然而，这仍然不可能被视为是一个足够的证据。这样的机器将会因为重复犯同样的错误，且不可能自我纠错或者只能依赖于外界干涉（来纠错）而大现原形。

如果一台机器能以某种方式“根据经验学习”，那将会给人以深刻的印象。因为如果这是真的，那么看起来我们没有任何理由不从一个相对而言较为简单的机器来开始这一切，通过让其经历一系列所谓的“经验”，逐步将其转化为一套精巧得多的机器，而且还能够被用来处理更为广泛的各类偶发事件。这个过程有可能以通过对其所需遵从的经验进行适当的选择而被加速，这可以被称为“教育”。但这里我们必须得谨慎从事。因为自动地使机器的结构以名为迭代、实为套用以前刻意设定的特定模式来安排相应的经验是十分容易做到的。显而易见，这就是一种彻头彻尾的作弊，因为这几乎等同于安排一个人躲在机器里面。这里再重复一下，这里所采用的关于认定何种教育内容是合理的标准是不能够随便套用所谓的数学体系的。但是我认为如下的陈述在实践中被证明是足够有用的：假定这个机器懂英文，由于它既没有手脚，也不用吃饭，更没有烟瘾，它会把时间主要用于下国际象棋和围棋，甚至桥牌。这个机器拥有一个键盘，任何想对它说的话都可以通过键盘被键入，同时它也可以输出任何它想说的话。针对这个机器的教育，我认为其应该被交给一些能够胜任且同时也对该项目感兴趣的教师。当然，该教师会被禁止获得关于这台机器内部功能的任何相关细节。建造这台机器的技术人员，可以被安排来保证这个机器的正常运行。如果他怀疑机器的运行不正常，他可以将机器设定回任何一种其先前的状态，同时让教师再从那一点开始，重新授课，但他本人并不参与整个的授课过程。由于这个步骤只是用来检验这个技术人员的善意与否，几乎不用说，在以后真正的实验阶段里，它并不会被采纳。实际上，在我看来，这个教育过程将会成为在合理的较短时间内生产出一台具备合理智能的机器的至关重要的部分。与人类自身所进行的类比也刚好说明这一点。

现在，我可以给出一些关于这样一台机器在未来将会以何种方式进行运行的特征描述。首先，这台机器会有一个内存，这一点并不需要做十分复杂详尽的解释。它应该是一个针对所有的陈述的简单的列表，包括为它而作的陈述以及由它而作的陈述，还有在游戏中它采取的所有的行动，以及它曾出过的所有的牌。这个列表将会按照时间先后来进行排序。在这些简单的存储之外，还会有一些“基于经验的索引”。为了解释这一理念，我将引入一个这种索引可能会用到的表格。它是一个针对被用到的词汇按照首字母进行排序的，同时还能给出其被使用的次数的索引表格。由此，它们可以在内存中被查询。另外，一个这样类型的索引可以包含人员的类型，或者对弈中的围棋盘面的几个部分。在对机器进行教育的较晚阶段，内存可以被扩展，用以记录在每一个时刻机器配置的重要零件，换句话说，机器将开始记住它曾经的想法。这将会产生出更多的新型的索引方式。在一些已使用过的索引中所观察到的特征指标意味着可以引入新型的索引方式。这些索引模式将会以下面这些方式被使用：无论何时，在要做出一个关于下一步将要做什么的选则时，当前状态的特征值会在已有的索引系统中被查询，在这种情况下，以前其在类似情况下所做过的选择，连同其相应结果，无论好坏与否，都会被发现。同时，相应的，新的选择也被制定出来。这就提出了一些新的问题：既如果一些指标是有利的，而另外一些指标是不利的，该如何是好？关于这一点，不同的机器会有不同的答案，同时，答案也会随其受教育程度的不同而有所变化。在最初阶段，可能一些十分粗糙的规则就足够了，例如，去统计谁获得来自支持者的选票最多等等。在整个教育过程相当靠后的阶段，机器自身可能会察觉到在这种情形下整个的关于步骤的问题，通过采用一些索引方式，可以产生出一些高度复杂的，而且如同其所希望的那种令人非常满意的规则形式。这看起来是完全可能的，尽管这些规则形式显得相对比较粗糙，但其自身就是相当令人满意的。所以，总的来说，尽管所选择的规则显得十分粗糙，但那个进程仍然可以被执行。这一点可以被一些看起来很复杂的工程问题有时居然是被那种基于经验的最粗糙的方法所解决而得到证明，而那些经验通常只是被用来处理那些问题最表层的部分。举例来说，函数是否随着它的一个变量的变化而增减。关于“决定行为”的方式所对应的的图景，其所提出的另外一个问题是“有利结果”的理念。如果没有这样一些和心理学家所倡导的“快乐原则”相对应的理念，继续向前迈进将会是非常的困难。所以，向机器引入类似的理念是非常自然的事情。我建议教师应该熟练使用两种关键的方法诀窍，其分别代表快乐与疼痛两种理念。在教育的较晚阶段，机器会识别出一些特定的其他的真切的情景感受，比如渴望，这由于它们曾经在过去被密切地经常性地与快乐联系在一起，当然也有其他的一些类似的理念，比如不被需要等等。甚至来说，来自教师的特定的生气表情也会被机器识别为坏事将近，而且这些表情不会被机器忽略，因此，教师会发现他也没必要再使用教鞭来督促机器的学习了。

在现阶段，沿着现有方向继续提出建议不会有太多效果，因为它除了对教育孩子所使用的真实的手段进行分析之外，极可能再没有别的有价值的东西了。尽管如此，还有一个特色，我想把它引入到机器里，这就是“随机要素”，每台机器将被提供一个磁带，里面有一个随机的数字序列，例如，相等数量的0和1，这个数字序列会被用于机器所做的选择，这会导致机器的行为无论以任何一种形式，都不会完全取决于其所服从的经验，而且，在人们和它互动实验时，这一点还会有别的有价值的用途。在某种程度上，通过制定欺骗性的选择，我们可以控制机器的发展方向。比如，人们可以坚持其所做的一个特定的选择，比方说，10个特定的数位（十位数？），这意味着在每1024台或者更多的机器当中，大约会有一台机器的智慧程度会发展到与被欺骗的那台机器一样的程度。由于“发展程度”这个理念所具备的主观本质属性，对这一点，我不能完美地给出一个很精确的描述，更不用说还有这样一个可能的事实：也许那台被欺骗的机器在其不被欺骗的情况下，其所做选择，已经是足够幸运了。

现在让我们假定，为了论据的充实，这些机器在未来的出现将会是一个切实的可能，然后看一下建造他们的后果。尽管自伽利略时代起我们已经在宗教容忍方面取得如此巨大的进步，但是这样做当然会招致来自部分人的强烈地反对。在反对的人们当中，肯定会有来自知识分子群体的强烈的抵触，因为他们害怕这样的机器肯定会导致他们失业。虽然关于这一点，知识分子们的认知可能是错误的，但实际上，这的确是可能的。当然，我们还有相当的工作需要去做，来尽力了解机器所要试图表达的意思，比如说，如果在未来，人们居然需要努力方能让自己的智能达到机器的水准，这样的话，似乎一旦机器的思维方法得到许可，被允许正式开始独立使用，其超越我们那微弱的智力是指日可待的事情，所以，这一切看起来是可能的，并非都是危言耸听。而且，机器本身不会有死亡的问题，它们也会通过相互交流学习来增进它们的智慧。所以在未来的某个阶段，我们不得不期望机器采取相应手段进行自我控制，比如采用前面提到过的Samuel Butler在其著作《Erewhon》所提及的方法。