

# 走向真正的人工智能

■ 张钹



张钹  
中国科学院院士，现任清华大学人工智能研究院院长。

我们现在离真正的人工智能还有一段很长的路。为了认清这个事实，需要回答下面三个问题：

第一，什么叫做真正的人工智能？我们的目标是什么？

第二，为什么我们需要真正的人工智能？

第三，我们如何走向真正的人工智能？

首先我们如何评价目前人工智能取得的成果，针对这五件事：

第一是深蓝打败人类国际象棋冠军；第二是 IBM 在电视知识竞赛中打败了美国的前两个冠军。这两件事是一种类型，后面的三件事是另外一种类型：即 2015 年微软在 ImageNet 上做图象识别，它的误识率略低于人类。百度、讯飞也都宣布在单句的中文语音识别上，误识率略低于人类。还有一个是 AlphaGo 打败李世石。

这五件事情都是机器在一定的范围内超过了人类。大家认为它们之所以成功，是由于三个因素：一是大数据，二是计算能力提高，第三是有非常好的人工智能算法。但还有一个因素是被大家所忽略的：所有的成果必须建立在一个合适的

应用场景下。

这五件事虽然领域很不一样，但是它们都满足下面五个限制：必须有丰富的数据或知识、确定性信息、完全信息、静态的、单任务和有限领域。任何一个条件不满足，现在的人工智能做起来都会非常困难。

这五个限制条件下的应用场景就是照章办事，不需要任何灵活性，这显然不是智能的核心。

例如下象棋是完全信息博弈，信息完全和确定。其次，它遵循着完全确定的游戏规则演化，我们把这种情况也叫做静态。Watson 机器人也是这样，知识竞赛提的问题都没有二义性，都是明确的，答案总是唯一性的。这样的问答对机器人来讲非常容易。它涉及的领域虽然比较宽，但也是有限的。包括大家觉得很玄的围棋，也完全符合上面五个条件，所以对计算机来说也是很容易的。目前计算机打麻将就不行，因为牌类是不完全信息博弈，所以比棋类要难。总之，我们对目前人工智能取得的成果要有一个正确的评价。

目前在交通、服务、教育、娱乐领域都可以找到人工智能技术的应用。但这些领域里只有满足上述五个条件的事情，计算机做起来才会容易。满足这五个条件的工作，总有一天会被计算机取代。就是那些照章办事，不需要任何灵活性的工作，比如出纳员、收银员等等。老师、企业家等工作不可能被计算机完全代替。

为什么有这五个限制？原因在于我们现在的人工智能是没有理解的人工智能。

先看符号模型，理性行为的模型。举 Watson 的例子，它是个对话系统，有知识库，有推理机制。除了专家知识之外，还有大量互联网上大众的知识，还运用了多推理机制，包括百科全书、有线新闻、文学作品等等。所有的知识用纸质来表示有两亿页，用存储量表示达到了 4TB。它能回答哪

些问题呢？用它的例子来说明。第一个问题，1974 年 9 月 8 日谁被总统赦免？用关键字“1974 年 9 月 8 日”“被总统赦免”，就能在文献里查出来是谁，答案是尼克松。根据问题中的关键字，可以在已有的文献里直接找到答案，这就是一般的网络检索方法。

第二个问题，荧光粉受到电子撞击以后，它的电磁能以什么方式释放出来？用“荧光粉”“电子撞击”“释放电磁能”等关键词，也可以找到答案：光或者光子。这种方法就是平时网络搜索的原理，应该说没有什么智能。

下面的问题就需要“智能”了：跟智利陆地边界最长的是哪个国家？

跟智利有陆地边界的国家可以检索到，它们是阿根廷和玻利维亚，但是谁的边境长？通常查不到。Watson 具备一定的推理能力，它从边界间发生的事件、边界的地理位置等等，经过分析推理以后可以找出答案，就是阿根廷。

下一个问题也属于这种性质：跟美国没有外交关系的国家中哪个最靠北？跟美国没有外交关系的国家有四个，只要检索就行了。哪个国家最靠北没有直接答案，但可以从其它信息中推导出来，比如各个国家所处的纬度、气候寒冷的程度等，分析出来答案是北朝鲜。

智能体现在推理能力上。但是很不幸，现在的对话系统推理能力都很差。Watson 系统好一些，但也很有限。

换句话说，现在的对话系统离真正的智能还很远。

我们通过索菲亚机器人可以看出来。索菲亚的对话是面向开放领域，你可以随便提问，问题就暴露出来了。大家在电视上看到索菲亚侃侃而谈，什么问题都能答得很好，这里面有玄机，如果你的问题是预先提出过的，因为里头有答案，就会回答得非常好，在电视上给大家演示的都是这种情况。如果我们临时提问，问题就出来了。这是一个中国记者给索菲亚提的四个问题，它只答对了一个。

“你几岁了？”

## THE FIVE LIMITATIONS

- Having rich data or knowledge
- Certain information
- Perfect information
- "Static"
- Single task and finite domain

These basically belong to well-defined routine jobs. They have a distance from the real AI.

目前人工智能取得成果的五个限制条件

这个问题很简单，但它答不上来，它的回答是：“你好，你看起来不错。”答非所问，因为它不理解你所问的问题。

只有第二个问题“你的老板是谁？”它是有准备的，所以答得很好。

第三个问题：“你能回答多少问题呢？”它说：“请继续。”没听懂！

再问第四个问题：“你希望我问你什么问题呢？”它说：“你经常在北京做户外活动吗？”这告诉我们，现代的问答系统基本上没有理解，只有少数有少量的理解，Watson 算是比较好的。

为什么会这样？我们现在的人工智能基本方法有缺陷，必须走向具有理解的 AI，才是真正的人工智能。这里提出的概念跟强人工智能有什么区别？首先我们都试图去准确地描述人类的智能行为，希望人工智能跟人类的智能相近，这也是强人工智能的一个目标。但是强人工智能只是从概念上提出来，并没有从方法上提出怎么解决。大家知道强人工智能提出了一个最主要的概念，就是通用人工智能。怎么个通用法？它没有回答。我们现在提出来的有理解的人工智能是可操作的，不只是概念，这是我们跟强人工智能的区别。

人机对话时，机器为什么不能理解人们提的问题，我们看一个例子就知道了。我们在知识库把“特朗普是美国总统”这个事实，用“特朗普 - 总统 - 美国”这三元组存在计算机里，如果你提的问题是“谁是美国总统？”机器马上回答出来：“特朗普。”但是你如果问其它有关的问题，如“特朗普是一个人吗？”“特朗普是一个美国人吗？”“美国有没有总统？”它都回答不了。它太傻了，任何一个小学生，你只要告诉他特朗普是美国总统，后面这几个问题他们绝对回答得出来。

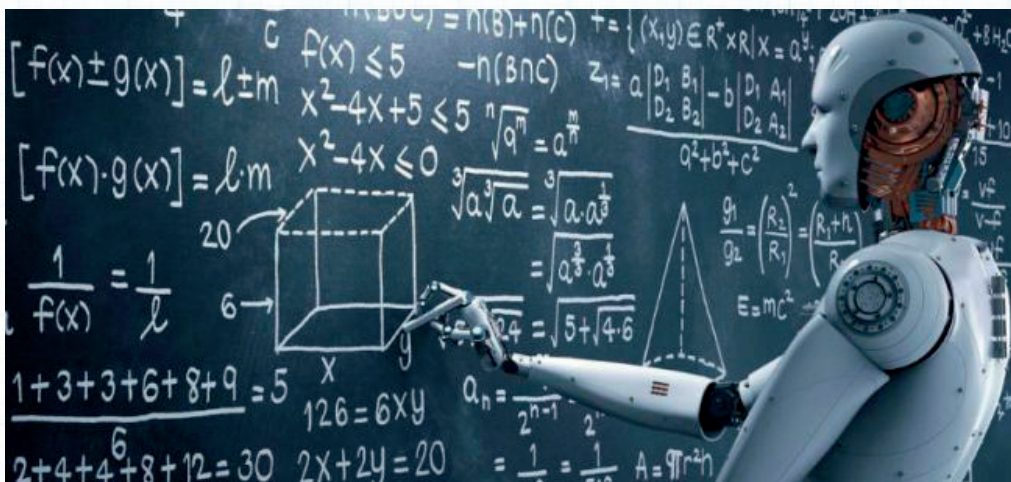
机器为什么回答不了后面的三个问题？就是

因为这个系统太笨了，没有常识，也没有常识推理。既然特朗普是美国的总统，美国当然有总统，但是它连这一点常识的推理能力都没有。要解决这个问题，必须在系统中加上常识库、常识推理，没有做到这一步，人机对话系统中机器不可能具有理解能力。但是大家知道，建立常识库是一项“AI 的曼哈顿工程”。怎么告诉计算机什么叫吃饭，什么叫睡觉，什么叫睡不着觉，什么叫做梦？这些对人工智能来说都非常难。美国在 1984 年就搞了这样一个常识库的工程，到现在还没完全做出来。可见，要走向真正的人工智能，有理解的人工智能，是一条很漫长的路。

这里介绍一点我们现在做的工作，我们的基本做法是建立一个常识图谱，用这个图谱帮助理解提出的“问题”，同时利用常识图谱帮助产生合适的答案。有了常识以后，对话系统的性能有了显著的改善，对话的质量提高了。

另外是准符号模型，深度学习、神经网络主要用来模拟感性行为，感性行为一般很难采用符号模型，因为感性（感觉）没法精确描述。比如“马”，怎么告诉计算机什么叫做马？马有四条腿，那什么叫做腿？细长的叫做腿，什么叫细？什么叫长？没法告诉机器，就不能用符号模型。目前用的办法是神经网络或者准符号模型，也就是人类的办法，学习、训练。不告诉机器什么叫做马，只是给它看不同的马的图片进行训练。训练完再用没见过的马的图片给它看，说对了，就是识别正确，说不对就是识别不正确。如果 90% 是对的，就说明它的识别率是 90%。

后来从浅层的神经网络发展到多层的神经网络，从浅层发展到多层有两个本质性的变化，一个是输入，深层网络一般不用人工选择的特征，用原始数据就行，所以深度学习的应用门槛降低了，你不用有专业知识，把原始数据输进去就行。



第二个是它的性能提高很多，所以现在深度学习用得更多，原因就在这里。

通过数据驱动建立的系统能不能算是有智能呢？必须打一个很大的问号。人脸识别系统甚至识别率比人还高，但是我们还不能说它有智能。这种通过数据驱动做出来的系统性能跟人类差别非常大，鲁棒性很差，很容易受干扰。

给定一个图像库可以做到机器的识别率比人还高，但是这样的系统只是一个机械的分类器，根本不是感知系统。它尽管把各种各样动物分得很清楚，但是它不认识这个动物。它只到达了感觉的水平，并没有达到感知的水平，只是“感”，没有上升到“知”。

我们的结论是，只依靠深度学习很难到达真正的智能。这是很严峻的结论，因为如果有这样的问题，在决策系统里不能用这样的系统，因为它会犯大错。人类的最大的优点是“小错不断、大错不犯”，机器最大的缺点是“小错不犯，一犯就犯大错”。这在决策系统里是不允许的，这就显示人跟机器的截然不同。人非常聪明，做什么事都很灵活，这使得他很容易犯各种各样的小错。但是他很理性，很难发生大错。计算机很笨，

但是很认真，小错误绝对不会犯，一犯就是天大的错误。人类只会把骡看成驴，但是计算机的识别系统会把驴看成一块石头。原因还是 AI 的理解能力问题。

自动驾驶过去讲得很多，我们看看问题在什么地方。现在我们通过数据驱动的学习方法，学习不同场景下的图象分割，并判别是车辆还是行人、道路等，然后建立三维模型，在三维模型上规划行驶路径。现在用硬件已经可以做到实时，如果路况比较简单，行人、车辆很少，勉强可以用。复杂的路况就用不了。原因非常简单，行人或者司机都会有意无意破坏交通规则，这就使得数据驱动方法失效。比如我们可以用数据驱动方法来了解各种各样行人的行为，进行大量训练，但训练完以后如果出现新的情况呢？自动驾驶不可能对付突发事件，如果这个突发事件它没见过，就解决不了。

怎么解决这个问题呢？实际上就是要解决从“Without”到“With”理解的问题。人工智能现在有两种基本方法，一种是用符号模型来模拟理性行为，符号模型可以表达信息的内容，所以它



张钹团队师生合影

是在一个语义的符号空间里，但是非常不幸，这个离散的符号表示，数学工具很难用，很多数学工具用不上去，所以它发展很慢。在模拟感性行为时，我们用的是特征空间的向量，向量就是数，可以把所有的数学工具都用上。所以数据驱动方法这几年发展非常快，再难的问题——下围棋非常难——计算机也可以“算”出来。但是它有一个非常大的缺陷，它是在特征空间里，缺乏语义。我们用数据去训练一个模型，所谓“黑箱学习法”，加上数据质量不高，很难学出有用的东西。什么叫概率统计？重复多了就是真理。如果数据质量差，充满了“谎言”，谎言重复多了，就变成真理了。

我们现在想出的解决办法是这样的，把这两个空间投射到一个空间去，这个空间叫做语义的向量空间。也就是说我们把符号变成向量，同时

把特征空间的向量变成语义空间的向量。一是通过 Embedding（嵌入）把符号变成向量，尽量保持语义不变，可惜现在的方法都会引起语义的丢失，我们只能在投射的过程中让语义丢失得少。第二方面做的工作比较少，就是 Raising（提升），把特征空间提升到语义空间去，这主要靠学科交叉，靠跟神经科学的结合。这些问题解决以后，我们才能够建立一个统一的理论。因为过去的感知和认知是不同的处理方法，大家说不到一块，如果我们能够投射到同一空间去，就可以建立一个统一的理论框架，这是我们的目标。但是这项工作是非常艰巨的。

还有一个办法就是把数据驱动跟知识驱动结合起来。人的智能没法通过单纯的大数据学习获得，那就加上知识，让它有推理、做决策的能力，这样就能解决突发事件。我们现在做

的工作就是把这些结合起来，知识也好，数据也好，都投射到同一空间，然后都用同样的数学方法进行处理。

最后做一个总结，我们从这个坐标（图1）看人工智能，横轴代表领域的宽窄，从单领域到多领域、到开放领域。纵轴代表信息的确信性与完全性，从完全到不完全、从确定到不确定。左下角代表最容易的部分，就是刚才讲的符合五个条件的，现在人工智能在这部分解决得非常好，我们用白色来表示它，AlphaGo 在这里，深蓝在这里，工业机器人在这里。现在我们正在向灰色地区去走，现在打德州扑克，一人对一人，计算机能战胜人类，多人对弈，计算机还不行，这是灰色地带，我们还可以做，尽管打牌是不确定的，但是它在概率意义下是确定的，你拿的这副牌的概率，同花的概率是多少，排成顺的概率是多少，可以算出来。既然概率能算出来，最终人类肯定会被计算机打败。

Watson 在右边，它的领域比较宽，但是它是确定性的，所以是在灰色的区域。往右上方去就比较难了，自动驾驶、服务机器人、大数据分析，它是一个大框，有的简单，有的困难。就自动驾

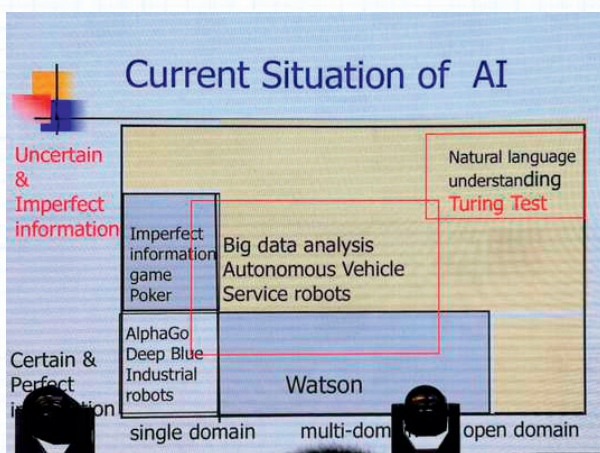


图 1

驶来讲，专用道、行车很少，路况简单等，在白色或者灰色区。如果路况复杂就到了黄色区域，黄色区现在计算机还解决不好。

最远的在右上角，图灵测试。图灵测试是开领域问答，很难！自然语言理解也在这里，复杂环境下的决策在偏左一点的地方，这也是很难的。所以人工智能现在是从左下角往右上角走，我们现在处在出发点附近。人工智能永远在路上，这就是人工智能的魅力，吸引我们去解决这些问题，一旦解决了，人类的生活就会发生本质上的改变。

最后我用一段古文作为总结：

周穆王西巡狩，路遇匠人名偃师。翌日偃师谒见王，偕来一个假人。“趋步俯仰，信人也”。领其颅，则歌合律；捧其手，则舞应节。千变万化，惟意所适。王以为实人也，与盛姬内御并观之，技将终，倡者瞬其目而招王之左右侍妾。王大怒，要杀这个偃师。偃师大惧，立剖其倡者以示王，皆傅会革、木、胶、漆、白、黑、丹、青之所为。穆王始悦，诏贰车载之以归。

这是 3000 年前古人对机器人的想象，现在的人工智能做得怎么样呢？索菲亚是我们现在达到的水平，可是她不会唱歌、不会跳舞，只会说英文。假设索菲亚“瞬其目而招王”，向大王送去秋波，王会不会大悦，神魂颠倒坠入爱河？我认为不会，因为索菲亚根本不像人，它最近才刚刚安上手脚，走路都不利索。所以我的结论是，“索菲亚通不过穆王的测试，当然它更通不过图灵测试”。

人工智能刚刚起步，离真正的 AI 还很遥远，大家共同努力吧，我们任重道远。☞

（本文为张钹院士为 CCF-GAIR 2018 主会场“AI 前沿技术”所做报告。）