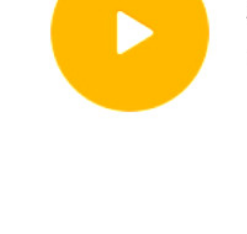


第150封信 | 等价问题和人工智能



吴军

第150封信 | 等价问题和人工..
11:05 5.19MB

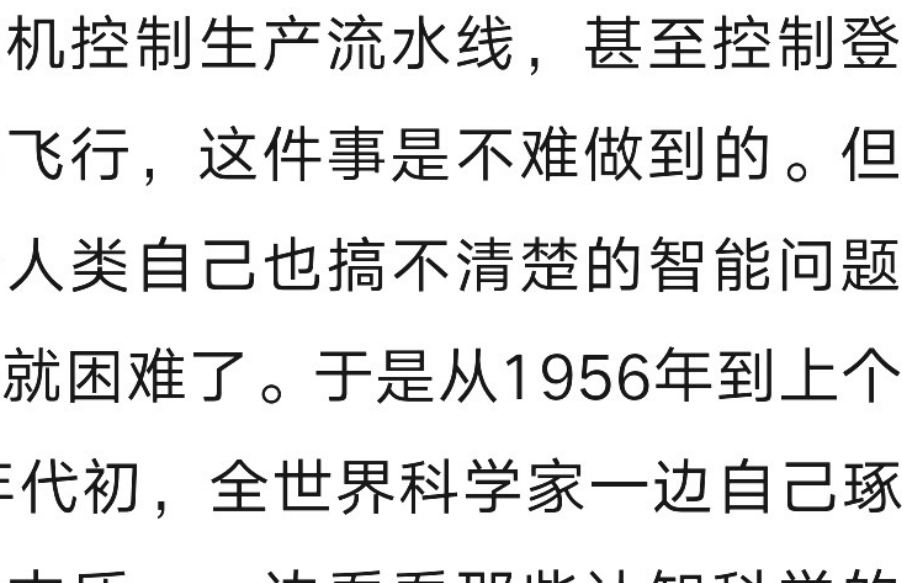
信件朗读者：宝木

小师弟，你好！

前一阵子因为中兴的事件，一家美国的公司也在媒体上经常曝光，它就是为中兴等很多通信公司提供芯片的高通公司。这周我们围绕高通公司的创始人维特比聊一些计算机思维的问题。不过今天我们还聊不到维特比这位科学大神和工业巨子，他会是我们明后天话题的中心，我们今天会介绍人工智能中的一大类问题，包括语音识别、机器翻译、自然语言的理解、基因识别和股票分析等等，背后的本质是什么。

我在今年的专栏和《硅谷来信》，以及很多公开的场合中讲过，人工智能发展的关键是找到相应的“空气动力学”。也就是说，正如我们无法依靠模拟鸟的飞行让飞机上天，而要通过掌握空气动力学的原理设计出能够有效飞行的飞机一样，人工智能也不是简单模仿人的思维和动作，而要找到适合计算机获取智能的“空气动力学”。

今天，人工智能的支柱其实有三个，就是摩尔定律、大数据和数学模型。我常常喜欢用下面一张图来表示上述几个概念和人工智能（也叫机器智能）的关系。



摩尔定律和大数据我们已经讲了很多次，这里不再重复了，我们今天从思维方式的角度来谈谈数学模型。

我们知道，要想让只会傻算题的计算机解决智能问题，就要找到一个相应的数学模型。对于一些确定性的问题，比如让计算机控制生产流水线，甚至控制登月火箭的飞行，这件事是不难做到的。但是，对于人类自己也搞不清楚的智能问题，这件事就困难了。于是从1956年到上个世纪70年代初，全世界科学家一边自己琢磨智能的本质，一边看看那些认知科学的方法是否可以量化地表述，从而使用计算机解决，大家在黑暗里探索了十多年，也没有取得什么令人满意的结果。

最先在人工智能的一个小领域取得突破性成就的是我多次提到的贾里尼克。他天才地找到了语音识别和机器翻译的等价问题——通信问题，从而通过解决通信问题解决了上述两个智能问题。

为什么说语音识别（包括对语意的理解）是一个通信问题呢？我们通常说到识别人的语音，就会想到声波、元音辅音（就是汉语拼音里的声母韵母）、词组、语法、上下文、完整的句子等声学 and 语言学的概念，其实这些都是表象，或者说是我们人为创造出来的概念。

过去科学家们就是这么想的，并且为每一个概念搞了一堆模型。但是在这么一大堆概念中找到一条解决方法并不容易。那么语音识别的本质是什么呢？我们不妨回到原点来想这个问题。我们通过说话，即口述语言的表达，是为了让收听者，明白我们脑子里想的意思。这个过程可以分成下面三个阶段：

1. 说话人将脑子里的意思用语言文字描述出来。
2. 文字变成发音，声音在空气中或者电话线中（包括移动通信的媒介中）传播，在收听方将声波变成我们耳朵能够接受的信号，通过听觉神经传到收听人的脑子里。
3. 收听人将收听到的信号还原回讲话人的意思。

上述的第一个过程，其实和通信中的信息源编码过程是一致的，你在通信中发一个邮件（包括微信和短信）也好，发一个照片或者视频也好，计算机先要将它变成能够在通信系统中传播的编码，这就是编码过程。

上述第二个过程虽然复杂，其实都是信息在不同信道中传输的过程。

而第三个过程，则是第一个过程的逆过程，即解码过程。

编码 - 信道传输 - 解码，这不就是一个标准的通信用过程么？既然语音识别的等价问题是通信问题，那么就可以采用解决通信问题的方法解决语音识别问题。看出了这一点，就是贾里尼克的过人之处了。

从1946年香农提出信息论之后，经典的通信问题都已经得到了解决。贾里尼克采用已知的通信模型，解决未知的等价问题，即语音识别问题，一举获得了成功。在贾里尼克之前，全世界努力了十几年，只能识别几百个英语单词，而且错误率还高达30%以上，贾里尼克带了几十个科学家干了五六年，就能识别22000个英语单词了，而且把错误率降低到10%。

当然，由于采用通信的模型需要大量的数据来训练模型，这种方法也被称为数据驱动的方法。

再接下来，贾里尼克的一个叫做彼得·布朗的部下发现机器翻译的问题也是一个通信问题。布朗是这么考虑的。假如我们要将中文翻译成英文，我们通常认为说话人说的是中文。但是布朗换了一个角度看问题，他觉得当你对于一个外国人说话时，说的是一种可以用英文表达的意思，只不过你在表达这个意思的时候，选择使用中文做了一次特殊的编码而已。翻译者，也就是解码的一方，所要做的无非是把中文编码的信息串解码还原成你的（英文）意思。这么看待问题，就把机器翻译问题从两种自然语言的理解，变成了一个通信问题。

我们可以把贾里尼克和布朗的思想，用这样一个公式概括：

语音识别≡通信问题≡机器翻译。

其中三个横杠，代表数学上的等价。我在之前不少封信中已经讲过，科学和工程的艺术常常体现在善于找到未知复杂问题的已知等价问题。这一点应该成为我们工作的方法论，很多时候我们所说的一通百通，其实就是在解决了一个问题后，知道经验可以往哪些问题推广。

至于为什么贾里尼克能想到语音识别这一类的智能问题是通信问题，这主要和他之前的经历以及专业知识储备有关。贾里尼克从来就不是传统的人工智能专家，因此在上个世纪60、70年代时，那群人的想法根本影响不了他，这让他避免走了弯路。但是，另一方面，作为香农的学生，他是那个年代最好的通信专家，并且写了当时大学信息论的教科书。从60年代初，他就在思考利用信息论解决语言问题，只是在70年代初，他到了IBM，有机会、有资源实现自己的想法而已。关于这段历史，如果你有兴趣，可以读我写的《智能时代》。

再接下来，是否还能找到这一类问题的等价问题呢？当然可以，到了90年代的时候，约翰·霍普金斯大学的萨尔兹伯格教授采用了一个全新的视角来看待基因测序问题。他认为，人的基因不过是一本特殊的书，书中的字母只有A、G、C、T这四种，因此任何识别语言文字的算法，比如语音识别和OCR的算法，都可以用于基因测序。

有了这个想法后，他离开了大学，到专门从事基因测序的研究所TIGR，转行从事基因测序研究了，并且后来获得了那个领域的最高奖。关于萨尔兹伯格教授的故事，我之前在硅谷来信第008封信中介绍过（末尾直达<硅谷来信>）。萨尔兹伯格和贾里尼克是大学的同事，他的想法当然是受到后者的启发，但是他能够最先看到基因测序和语音识别，以及和通信是等价问题，并且以此为钥匙解决问题，这是他的过人之处。

这一类的等价问题还可以举出很多。简单地讲，它们可以概括为两类：第一类是直接等价问题，包括股票市场的分析和预测；第二类是间接的、广义上的等价问题，包括今天很多图像识别，特别是人脸和医学影像识别的问题。

在解决第一类的问题中，前面提到的布朗做了很好的尝试并且获得了巨大的成功。在IBM从事机器翻译时，由于缺乏数据量，因此布朗的研究结果其实不算太好，他那篇划时代的论文在开始的十多年甚至没有什么人引用。当然，布朗可以等个十几年，这显然不高明，于是他跑到一个有数据的地方去做一件别的事情了。他去的地方就是我在过去提到的，近30年来