人工智能的尽头是人工（二）偏差与方差之争

作者：张军平 来源：科学网博客

本文原地址：https://www.iikx.com/news/topnews/11081.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

人工智能的尽头是人工（二）：偏差与方差之争。自上世纪50年代开始，人工智能研究关心的核心问题之一就是预测。如早期的跳棋程序设计，就期望能获得对棋局的准确预测。但当年人工智能的学习算法，在研究水平上还达不到实用级，数据的采集规模和条件都很有限，硬件条件也不允许做基于大规模数据的高效计算。最初曾考虑模拟人的决策方式，比如采用基于规则的方法或专家系统。但好景不长，不多久基于规则的方法和专家系统就面临组合爆炸问题，即无法穷尽所有的可能，总有不符合规则的例外出现。所以，这一思路没持续多长时间就被其它更有效的预测方法替代。

而同时期Rosenblatt提出的感知机模型，让大家看到了利用有限的样本数量，从理论上估计分类器性能的希望。但这一方法风光一段时间后，就被人工智能代表人物Marvin Minsky指出无法解决异或问题。神经网络方向的研究也因此一度陷入困境。再加上同时期其它人工智能研究的不顺利，使得人工智能进入了第一次寒冬。

虽然如此，人类在提高模型的预测能力上还是想了很多办法，复杂的、简单的，统计的、几何的。不过用哪种方法，事实上都离不开两个基本概念的平衡或折衷，即偏差和方差。

偏差是什么呢?就像有条正确的路，然后你走的时候因为重心不稳导致无法一直沿着它走，便产生了与正确路线不同的差异，即偏差。

方差是什么呢?哲学家赫拉克利特说过，一个人不可两次同时踏进同样的河。即然无法沿精确的正确路线走，那每次走的路线自然也会不一样。把多次行走出来的路线累计起来再算对这些路线的平均结果的差异，再平方后就是方差。

我们在预测时，往往希望偏差尽可能小，以便获得好的预测。同时，也希望方差小，这样模型会更稳定。然而，这两个却天生就有矛盾，如同熊掌与鱼，不可兼得。

偏差小的，通常需要对真实的路线做更精细的匹配，这就意味着得把模型设计得更为复杂才有可能。它带来的副作用是其对数据的敏感性会增大，预测的结果容易产生更大的波动，即大的方差。偏差大的，模型相对简单，比如不管是啥路，我就走直线，那从数学上看显然是最简单的，稳定性也好。所以，对不同的路线，结果产生的波动相对也要小，即小的方差。

从人工智能角度来看，要对真实任务建模并形成好的预测性能，往往是三部分的平衡所致。一是偏差，二是方差，三是不可约简的噪声。第三部分一般认为是固有的、无法消除，所以，对世界的学习，重点主要都放在前两部分。

即然两部分的和决定了对世界的逼近能力，自然就可以沿两条不同的思路来实现。

一种是尽可能是减少偏差。以分类为例，即识别一个目标属于哪一类，能最大程度减少偏差的似乎是最简单且几何上直观的1-近邻分类器。按字面理解就知道，它是根据离哪个已知标签或类别的样本更近来判别未知样本的类别归属。这种方法只要衡量远近的距离或度量确定了，后面的处理都简单易行，找到1个最近的即可。所以，只要有标签的训练样本足够多，可以保证偏差会很低。但问题是，这样做的话，容易一叶障目，导致看不全未知的变化，以至于方差会比较高。Cover和Hart 1967年曾给出过一个著名的结论，即从渐近意义来看，1近邻分类的错误率不超过贝叶斯误差率或"人类误差率"的两倍。粗略来说，就是最多训练数据和测试数据会各贡献一次误差。

另一种则是减少方差，它寻找对真实世界的稳定逼近或近似。举例来说，统计学比较喜欢用的最小二乘。该方法是通过简单的特征加权组合来实现对未知世界或函数的逼近。这一方法的好处是线性，能提供强的可解释性。由于是无偏的，即估计量的数学期望等于被估计参数的真实值，著名的高斯--马尔可夫定理曾证明过，在所有无偏的线性估计子中，最小二乘估计具有最小的均方偏差。

有了对偏差方差的直觉印象，科研人员发现其实在线性情况下做的无偏估计，有的时候总的偏差仍然不低。那么，如果不考虑无偏，稍微放松点，做成有偏的估计，虽然会增加一些方差，但却有可能进一步减少偏差，从而让预测性能能够进一步提升。

举例来说，偏最小二乘方法搜索有高方差和与响应高度相关的方向，并倾向于收缩掉低方差的方向。虽然不是无偏的，却也能进一步提高总的预测性能。类似地，岭回归是在所有方向均进行收缩，但对低方差方向收缩更为厉害。也有反向的处理。如自然三次样条，则是通过约束边界节点以外的部分为线性函数，以边界处偏差增大的代价来减小边界节点以外的方差。诸如此类的技巧还有很多，就不一一枚举。

在经典的偏差方差理念下，大家想到的解决方案或模型设计都是寻找这两者的平衡或折衷。然而，对实际工程应用中，事实上，有的时候也可以忽略统计上的期望，只追求对单次或有限次数意义下的有效估计。那么在这种情况下，我们可以找到更优异的模型，比如深度学习。

除了偏差与方差，统计与个体的思考外，还有科学家考虑过(非不可约)噪声存在的情况下，需要做的折衷。

比如提出过《控制论》、或者说自动化专业的“ 祖师爷 ”维纳。他发现信号的还原和噪声的抑制之间存在折衷。具体来说，在还原信号的过程中，如果只是单纯的把让信号退化的函数逆变回来，在有独立于信号退化过程的噪声存在时，那么这个噪声也得除下退化函数。结果是，当该退化函数在频率高的部分系数很小时，这些小的系数会在还原真实信号的同时放大噪声的不利影响，导致逆向还原函数的性能不理想，甚至有可能完全受噪声的影响。此时，就需要引入能自适应控制信噪比的维纳滤波器来处理。

不难看出，为了能获得好的预测性能，我们在偏差方差上花了非常多的时间和精力去寻找突破。但需要指出的是，人类和智能生命存在的意义可不止是为了预测。预测是帮助其生存的重要条件，但并非唯一意义。如果只是预测，人类和其它智能生命可能就只是机器了。这也是人工智能和人工的智能存在最大区别的地方。

张军平

2020年9月15日

注：在追求可解释性的前提下，科研工作者最爱用的工具是线性的，比如长宽高这种可以简单叠加计算的，就属于线性的。之所以线性意味着可解释，原因之一是因为多数人类只懂加减乘除。即使在大学中学过更为高深的高等数学，一旦进入社会，最终能保留下来的也就是利用excel电子表格来做简单的加减乘除运算了。所以，把模型做成线性，有利于我们理解和解释其中的道理，哈哈!

(原创作品，如需转载，请联系我本人，谢谢!)



张军平，复旦大学计算机科学技术学院，教授、博士生导师，中国自动化学会混合智能专委会副主任。主要研究方向包括人工智能、机器学习、图像处理、生物认证及智能交通。至今发表论文近100篇，其中IEEE Transactions系列20余篇，包括IEEE TPAMI, TNNLS, ToC, TITS, TAC, TIP等。学术谷歌引用4000余次，H指数32。

出版科普著作《爱犯错的智能体》，曾连续24次推荐至科学网头条，2019年获得中国自动化学会科普奖。关于人工智能发展趋势的观点曾被《国家治理》周刊、《瞭望》、《科技日报》、《中国科学报》等媒体多次报道。连载的《读研秘技》至今被科学网推荐头条20次。

更多 科研头条 请访问 https://www.iikx.com/news/topnews/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发