06 决策：AI的预测 + 人的判断

万维钢·AI前沿 2023年4月7日

06决策：AI的预测+人的判断.mp3
13分38秒

转述：怀沙

你好，这里是《万维钢·AI前沿》课。这一讲，咱们来说说人和AI的决策分工。

人到底有什么能力是不可被AI替代的？每个人都需要思考这个问题。

我们在第四讲说到三个加拿大经济学的一本书，《权力与预测：人工智能的颠覆性经济学》。这本书中有一个洞见，我认为有可能就是AI和人分工的指导原则。简单说，就是双方共同做出决策，其中AI负责预测，人负责判断。

要说明这一点，我们先看一个真实的案例。网约车公司Uber一直在测试自动驾驶汽车。2018年，Uber的自动驾驶汽车在亚利桑那州撞死了一个行人，引起了激烈的讨论。

你仔细分析这个事故，会发现在撞击前6秒，AI已经看到了前方有一个未知物体。它没有立即做出刹车的决定，因为它判断那个物体是人的概率非常低——虽然那个概率并不是0。

AI有个判断阈值，只有在前方物体是人的概率超过一定数值的情况下，它才会刹车。撞击前6秒，概率没有超过阈值；等到后来终于看清是人的时候，刹车已经晚了。

我们把刹车决定分为「预测（Prediction）」和「判断（judgement）」两步。AI的预测也许不够准，但是它已经预测出这个物体*可能*是一个人，它给出了不为0的概率——接下来的问题是出在了判断上。在这个概率上应不应该踩刹车，是这个判断导致了悲剧。

Uber AI 用的是阈值判断法，这可以理解，如果对前方任何一个是人的概率不为0的物体，AI都选择刹车，它就会在路上不停地踩刹车，这车就没法开了。当然你可以说这个阈值不合理，但是这里总是需要一个判断。

请注意，正因为现在有了AI，我们才可以做这样的分析。以前发生过那么多人类司机撞人的事件，从来没有人去分析那个司机是犯了预测错误还是判断错误。但这种分析其实是完全合理的，因为两种错误性质很不一样——请问这位司机，你是根本没看见前方有人呢，还是已经感觉到前方物体*有可能*是人，但是你感觉那个可能性并不是很大，又因为赶时间，你觉得那么小的概率可以接受，就开过去了？

你犯的到底是预测错误，还是判断错误？

*

决策 = 预测 + 判断。

预测，是告诉你发生各种结果的概率是多少；判断，是对于每一种结果，你在多大程度上愿意接受。

我在精英日课第五季，讲过蒂姆·帕尔默的《首要怀疑》，其中专门讲过怎样根据预测的概率做决策。你周末有个户外聚会，要不要为此租个帐篷防止下雨，这是你的决策。天气预报告诉你那天下雨的概率是30%，这是预测。面对这样一个概率，下雨的损失是不是可以接受的，这是你的判断。

我们当时讲的是只要采取行动的代价（也就是帐篷的租金）小于损失（也就是下雨会给你带来多大的麻烦）乘以概率，就*应该*采取行动，租个帐篷防止淋雨。但是在这一讲的视角下，请注意，这个“应该”，应该理解成是对你的*建议*。

是否采取行动的拍板权还是在你手里——因为那个损失最终是由你来承受的。AI不会承受损失，用公式给你提建议的人不会承受损失。英国女王也好、你岳母也好，在场来宾淋雨这件事是大是小，不是AI所能知道的——那其实是你自己的主观判断。

AI很擅长预测天气概率，但是判断一个天气状况带来的后果，需要更多的、具体的、也许只有你自己才知道的信息，所以做判断的应该是你而不是AI。

AI时代的决策 = AI的预测 + 人的判断

也就是说，我们应该让预测和判断脱钩。以前所有的决策都是人负责预测，人负责判断，现在则应该是AI负责预测，人负责判断——

我们承认AI比人聪明，但是我们也知道真正承受风险、体验后果的是人而不是AI。预测是客观的，判断是主观的。AI不能僭越人的判断，人也不应该专断AI的预测。AI与人各安其位，分工明确。

*

如何实施这个分工呢？一个方法是人为给AI设定一个自动判断门槛。

比如自动驾驶汽车，我们可以规定，当AI预测前方物体是人的概率高于0.01%——或者0.00001%也行，反正你得有个不为零的数值——的时候就必须踩刹车。这个判断标准，这条线，肯定不是AI自己规定的，而是人事先设定的。你把这条线编程到AI之中，但是下达那条编程指令的必须得是人——因为只有人能判断人命的价值：对AI来说，人命的价值是无法用客观方法估算的。

其实我们已经在用这种判断了。以前你到商店买东西用的是现金，那个现金是真钞还是假钞，得由商店收银员自己预测、自己判断。现在你刷信用卡，那个信用卡是真卡还是假卡，现在不是由收银员决策，而是由信用卡联网系统根据算法来预测和判断的。算法会先评估这张卡是假卡的概率有多大（预测），再看看那个概率是否高于某一条线（判断），决定是否拒收。那条线不是任何AI算出来的，而是事先某个由人类组成的委员会划定的——因为线划得太低得罪客户的是人，线划得太高承担损失的也是人。

未来我们会面对各种各样类似的事情，《权力与预测》这本书建议这样的判断最好由一个像美国食品药品监督管理局（FDA）这样的机构来执行，就好像评估一种新药是否可以上市一样。

*

另一个方法是把判断量化成钱。

你租了一辆车，要去一个比较远的地方，你有两条路线可选。第一条比较直，老老实实开车就行，路上没什么风景。第二条会经过一个风景区，对你来说是一种享受，但是风景区里有行人，会增加出事故的概率。如果AI直接跟你说哪条路出事故的概率有多大，你可能还是不好判断。

更方便的做法是，AI告诉你，走风景区那条路，租车的保险费比走第一条路高1块钱。这一块钱的保险费代表AI对两条路风险差异的预测。

现在判断交给你。如果你认为风景对你的重要性超过1块钱，那你就走风景区；如果你对风景没有那么高的兴趣，你就省下1块钱。

你看，AI无需了解你，也不可能了解你——是你在这一块钱和风景之间的选择，揭示了你的偏好。在经济学上，这叫做「显性偏好（Revealed Preference）」：人的很多偏好本来是说不清的，但是一跟钱挂钩就能说清了。

*

预测跟判断脱钩，对人是一种赋能。

以前如果你想去开出租车，可不是会开车就行。首先你得学认路，你得知道这个城市中从任意A点到任意B点的最短路线是什么——用本讲的话说就是你得会预测——才能开好出租车。现在AI接管了预测路线的事儿，你只要会开车就可以去开网约车了。

有了AI，人会判断就会决策。但这并不是说决策很容易，因为判断有判断的学问。

生活中更多的判断既不是由委员会划线也不是被量化成了钱，是必须由个人对具体情况进行具体分析。这个结果对你来说到底有多好，或者到底有多坏，到底能不能承受，你怎么判断呢？

有的可能是你读书或者跟别人学的，比如你听说被烧红的烙铁烫会很疼，你会愿意以很高的代价避免被烫。但是听说不如亲历。只有你真的被烫过，你才能知道有多疼。判断，有很大的主观成分。

而判断这个能力正在变得越来越重要。美国的一个统计显示，1960年只有5%的工作需要决策技能，到2015年已经有30%的工作需要决策技能，而且还都是高薪岗位。

因为只有人知道有多疼，所以人不是机器。而判断力和随之而来的决策力，本质上是一种权力——AI没有权力。这就是《权力与预测》这本书书名的缘起。

购买课程微信：15612777539

*

当AI接管了预测之后，决策权力在社会层面和公司组织层面的行使就成了一个新问题。

有个例子。因为知道了铅对人体有害，从1986年开始，美国政府禁止新建筑物使用含铅的饮用水管。可是很多旧建筑物的水管都含铅，这就有一个对旧水管的改造问题。但是改造非常费钱费力，旧水管又不都含铅，你得先把水管挖出来才知道是否含铅，那先挖谁家的呢？

2017年，密歇根大学的两个教授搞了一个AI，能以80%的准确率预测哪家的水管含铅。密歇根州的弗林特市（Flint）使用了这个AI。一开始都挺好，市政府安排施工队根据AI的预测给各个居民家换水管。

工程这样进行了一段时间之后，有些居民不干了。他们质疑说，为什么我家邻居水管换了，我家的却没换？特别有富裕居民区的人说，为什么先去换那些贫困地区的水管，难道不是我们交的税更多吗？

收到这么抱怨，弗林特市的市长就干脆决定不听AI的了，咱挨家挨户慢慢换。结果这样一来，决策准确率一下子从80%降到了15%……又过了一段时间，是美国法院推出一个法案，说换水管这个决策必须先听AI的预测，决策准确性才又提高回来——

这件事儿的道理是AI改变了决策权。没有AI的预测，因为只有政府能预测和判断谁家水管应该先换，决策权是完全把持在政客手里；有了AI，老百姓或者社区都可以自己判断，尤其美国三权分立，司法系统也可以直接发话，政客就不好使了。

*

决策权到底应该属于谁呢？从经济学角度来说，当然是谁决策对整个组织的效率最高，就应该属于谁。

以前预测和判断不分的时候，决策权往往应该交给一线人员，因为他们直接接触关键信息，他们的预测最准确——正所谓「让听得见炮火的人指挥」。

现在AI接管了预测，人的决策就是判断。这时候可以考虑，让那些个人利益跟公司利益最相关的人决策，或者让受这个决策影响最大的人决策，或者让最能理解决策后果的人决策……这些都意味着组织的变革。

变革还意味着以前的预测者，现在要转型为判断者，或者解释者。

比如以前天气预报机构的职责是提高预测的准确度，现在有了AI，他们的主要职责就不是预测天气了——也许变成了向公众和政府机构解释预测结果。AI说下周有5%的可能性会发生龙卷风，政府官员不懂这5%代表多大损失，也许你这个气象学家能给解释一下。未来的气象台可能更多地是提供人性化服务，比如建议老百姓明天怎么制定出行计划，而不仅仅是预报一个下雪概率。

再比如，以前放射科医生的最主要任务是看图预测病情。现在AI看图的能力已经超过了人类，那放射科就得琢磨别的服务，也许是向病人解释病情，也许放射科就不应该继续存在……

*

好，我们来总结一下这一讲。《权力与预测》这本书提供了一个解决方案，那就是AI只负责预测，让人掌握判断。在我看来这个分工非常合理，把决策权交给AI会让人很难受。AI再厉害，我们也不愿意把它奉为神灵——AI最好老老实实扮演助手或者祭司的角色，拍板权还是应该在人的手里。

以上就是这一讲的全部内容，下一讲咱们继续聊AI这个危险的法宝。

划重点 添加到笔记

AI时代的决策 = AI的预测 + 人的判断
预测是客观的，判断是主观的。AI不能僭越人的判断，人也不应该专断AI的预测。AI与人各安其位，分工明确。
实施这个分工，有两个方法：
1.人为给AI设定一个自动判断门槛。
2.把判断量化成钱。

首次发布: 2023年4月7日

我的留言

得到
写留言，与作者互动
0 / 5000
公开 发布

用户留言

最新 最热 只看作者回复

Skye 04-08 + 关注
AI 时代，以 AI 作为一种高级使用辅助工具，最后还是归于人和人的的判断，人的决策是万万不能不能被代替的，是人划了那条线，人受到也才懂得最终的影响。
新时代要努力转型去判断，去做决策。
转发 评论 8 分享

Sandwich 04-08 + 关注
归因于打了太多次数学建模比赛，预测和决策模型对我来说太亲切了。万老师讲的行人避让预测 - 决策模型，我脑海中就自主构建了一个模型。
模型前提：人命的价值在当代还是不可量化的，我们无论如何都要尽力避免车辆冲向行人并导致死亡的事件发生。对于减少这个现象的发生，我有两个优化的思路。
1. 提高预测的准确率，不断优化预测模型的硬件和软件，让预测的准确率持续提高，以此使汽车能够准确的识别行人。
2. 改进决策阈值，对于预测前方是行人的概率而进行避让的阈值不一定是一个单一的常数，而是一个随车速和路况变…
展开
转发 2 11 分享

爱学习的猫 04-08 + 关注
在 AI 元宇宙的未来幻想怎样？

在元宇宙中，不管以什么
意识形态出现，我们百姓
注定会变成某种 AI 世界的道具
就以目前的智能软件为例
它们背后的运行程序，有意地…
展开
GPT学习小组
转发 评论 6 分享

亚义 南京 04-07 + 关注
“AI 将要替代很多的工作”。
“AI 将要为我的工作赋能”。
前者是正确的静止，后者探索中有发展。
转发 评论 18 分享

祝韬 04-07 + 关注
我来补全公式。
分为「预测（Prediction）」和「判断（judgement）」两步
决策 = 预测 + 判断。
自己：判断 = 收益 * 收益概率 - 损失 * 损失概率
预测，是告诉你发生各种结果的概率是多少：判断，是对于每一种结果，你在多大程度上愿意接受。
AI 很擅长预测天气概率，但是判断一个天气状况带来的后果，需要更多的、具体的、也许只有你自己才知道的信息，…
展开
2 1 26 分享