
将决策权“外包”给AI，靠谱么？

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/14294.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！

将决策权“外包”给AI，靠谱么？。

从个人衣食住行到国际间邦交誓盟，选择或决策无处不在。

很多时候，选择或决策是困难的。人们不但会面临诸多两难选择，还经常受到情绪、偏好、认知水平等因素的影响。同样，精准地预测人类的选择也非易事，在人工智能（AI）和大数据相关技术日新月异的今天，将决策和预测决策结果外包给人工智能，也许是个不错的选择。

近日，上海大学悉尼工商学院（拟聘）副教授何黎胜和美国宾夕法尼亚大学Sudeep Bhatia合作，发表在《科学》的文章，分析了深层神经网络模拟人类行为特征，在预测人类决策领域起到的重要作用。

预测决策行为的瓶颈

以20%的概率得到100美元，或以80%的概率得到50美元。如果面临这样的选择，你会选择那个？

1979年，诺贝尔经济学奖获得者卡尼曼等人提出的前景理论（Prospect Theory）认为，人们在面临获得时往往小心翼翼，不愿冒风险；在面对失去时会很不甘心，更容易冒险。人们对损失的痛苦感要大大超过获得时的快乐感。

前景理论为人们理解决策者如何组织决策提供了模型，但用前景理论来模拟选择行为并非没有缺点。依据前景理论提出新理论的研究者通常会对诸如感知、注意力、记忆和情感等过程，以及干扰和选择错误的原因做出复杂的假设。

Sudeep Bhatia认为，这些理论本身只在小数据集的选择上进行测试，很少与大数据集的已有模型进行比较。考虑到决策研究的跨学科历史和风险选择的复杂性，这是不可避免的：决策者很容易凭直觉对预期效用的偏差做出心理解释。而且，很多新理论模型通常类似于先前发布的模型，许多理论在基准数据集上高度模仿彼此的预测。

尽管新理论模型产生的速度正在加快，但在过去20年中，这些数据集的预测精度几乎没有提高。何黎胜告诉《中国科学报》，其根本原因是我们的认知系统具有高度复杂性。

何黎胜解释说，风险决策看似简单，如果我们问一个大学生（或高中生）应该怎么作风险决策，大多数人可能会说我们先算出每个选项的期望价值，然后选择期望价值最高的选项。

但风险决策比这复杂得多，受到多种认知和情感因素影响。以往研究者通过考虑这些影响因素，对风险决策行为提出了多种理论解释模型。

近20年来，这种‘小数据集+小模型’的研究范式遇到了瓶颈，主要表现为我们对决策行为的解释越来越多，但对决策行为的预测能力并没有显著提升。何黎胜说这很大程度上取决于常用的‘小数据集’和‘小模型’（与深度神经网络相比，常规的决策模型都是小模型）研究范式。小数据集不足以支撑像深度神经网络这样的‘大模型’，而‘小模型’又不足以支撑决策过程中复杂的认知与情感因素。

大模型显身手

人工智能技术虽然发展迅速，但计算机不会有人类意义上的感情，短期内也不可能产生人类特征的情绪。天津大学自动化学院副教授杨正瓴对《中国科学报》说，这也许会成为人工智能的优势，让它在某些领域比人类做得更好。比如，利用‘大数据’的人工智能决策，的确可以帮助人类减少决策失误。

近年来，不同学科之间的交叉研究提供了新的突破口，特别是行为决策、认知科学和机器学习等领域的交叉。

总体而言，在‘小模型’范围内，前景理论的预测精度是上佳的，是迄今为止预测精度最好的风险决策模型之一。何黎胜说，如果我们把视野放到更广阔的‘大模型’（例如深度神经网络等机器学习算法），前景理论等小模型的预测劣势就显现出来了。

不久前，美国学者乔舒亚·彼得森（Joshua Peterson）等人发表于《科学》的文章，系统地比较了大模型与小模型（如前景理论）对风险决策行为的预测能力。结果发现，深度神经网络的预测能力比前景理论高出几个能级。研究者首先针对10000多个不同的选择问题（涉及概率货币回报的赌博，超过了先前数据集的大小）和人类最终在这些问题中做出的决定，来训练深度神经网络，发现这种网络能够以非常高的准确率模拟人类决策，大大优于现有的（人为）风险选择模型。

在我与南丹麦大学教授Pantelis Analytis、和宾夕法尼亚大学教授Sudeep Bhatia共同完成的另一项研究中，通过群体智慧算法整合不同的认知与情感因素。我们发现这样的集成模型能显著地提高对决策行为的预测能力。何黎胜说，这里的集成模型也是一种‘大模型’，它的预测精度也超过了前景理论。

前景理论提出至今，给后续研究带来极大的启发，成为决策科学中的经典研究。当前，研究人员在此基础上，利用人工智能和大数据算法，一方面不断完善前景理论，另一方面通过新的技术方法不断拓展决策理论的边界，提高预测精度。

实现决策智能化

经过学习和训练，深度神经网络能够模仿人类的行为，它甚至可以像人类在做选择时，不理性的行为那样，表现出非理性。Sudeep Bhatia告诉《中国科学报》，因此，用它来预测人类决策行为会大大提高预测精度。

人工智能、认知神经科学的发展，将为心理学、决策科学的发展提供有力的研究手段。何黎胜说，得益于学术界和工业界的大模型投入，近年来人工智能和认知神经科学领域发展迅速，涌现出

大量的开创性研究，也为其他学科提供了成熟的研究工具。

何黎胜认为，在决策研究中，我们会研究不同类型的决策，像风险决策、跨期决策、社会决策、博弈决策等。目前，对这些不同决策类型的研究和建模工作通常相互独立。然而从认知的角度，不同的决策类型必然有互闻共通的认知与情感机制。而不同决策类型之间互闻共通的建模研究还非常稀缺，未来完全有可能通过与人工智能、认知神经科学融合，实现不同决策类型的互闻建模。

机器生成理论在决策和管理场景中有巨大的应用潜力，其核心的变革在于管理自动化。这个转变类似于计算视觉研究中的深度神经网络模型在医学图像处理中的应用。何黎胜说。

日前，管理科学越来越趋向于数据驱动的决策。在工业界和公共政策等领域，越来越多的公司和组织实验方法（如A/B测试）获得实证数据，通过分析实证数据达到决策优化的目的。然而，目前这种数据驱动的决策有不小的局限性。其中一个重要原因是管理者很多时候面临的是一个复杂的动态系统，是由多种因素交织而成的综合体。但A/B测试等方法收集到的数据往往是针对某个特定的小点，对复杂决策中的多因素之间的交错互动缺乏全局把握，限制了A/B测试实验数据的适用性。

机器生成理论可以更大程度地利用A/B测试等方法收集到的数据。更为重要的是，机器生成理论足够灵活，我们可以用真实的数据去训练这个系统，让它不断的逼近真实的管理场景何黎胜说，未来，训练好的机器生成理论模型可以自动地对复杂管理场景中的事件做出反应。也就是说，机器生成理论有能力将分散的实验数据升华成具有逻辑一致性的体系，实现复杂管理决策的自动化、智能化。（来源：中国科学报 张双虎）

相关论文信息：<https://doi.org/10.1126/science.abi7668>

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

作者：何黎胜等 来源：《科学》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发