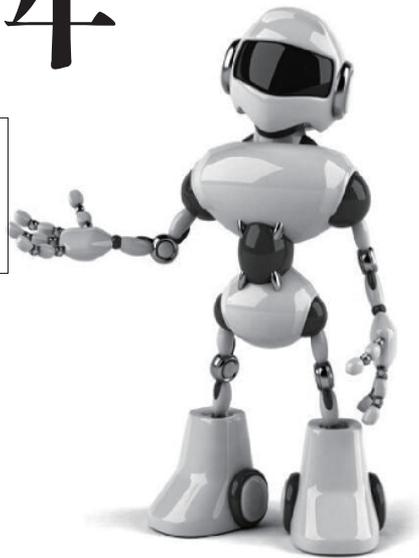


# 入选 MIT2018 十大突破性技术的 AI 版“双手互搏”有多牛

**核心提示:**一个网络中有两个角色,修炼的过程中左手扮演攻方,即生成器(generator),试图生成和自然世界中拟完成任务足够相似的目标;右手扮演守方,即判别器(discriminator),试图把这个假的、生成的目标和真实目标区分开来。经过反复多次双手互搏,左手右手的功力都会倍增,从而达到“舍我其谁”的目标。



日前,《麻省理工科技评论》刊文评出了 2018 年十大突破性技术,“对抗性神经网络”(GAN)赫然在列。

什么是对抗性神经网络?为什

么它能入选 MIT 十大突破性技术?它的发展脉络如何?与我们此前耳熟能详的神经网络有什么区别?能够应用在人工智能的哪些场景?还有哪些关键问题有待攻克?

中国自动化学会混合智能专委会副主任、中国人工智能学会机器学习专委会常委、复旦大学博士生导师张军平教授在接受记者采访时做了深入浅出的解释。

## 故事中的 GAN 幻影

张军平告诉记者,虽然 GAN 是科技领域的“新贵”,但其实在很久以前的中外小说中就能看到这个想法的影子。

这个想法最早可以追溯到奥地利小说家斯蒂芬·茨威格写于 1941 年的小说《象棋的故事》。

在该小说中,主人公 B 博士被长

期囚禁在纳粹集中营中。在想尽了各种摆脱空虚和孤独的办法却徒劳无功后,B 博士意外得到了一本国际象棋棋谱。

他背完书里的上千棋谱后,用送来的面包做了副国际象棋,开始自己跟自己下棋,最终演变成一盘接一盘的相互疯狂挑战,这令他棋力大涨。出狱

后,在一艘游艇上,他居然轻松地在首盘就击败当时的世界冠军。

在中文小说里也有类似的影子,它出现在金庸写于 1957 年的武侠小说《射雕英雄传》中。

王重阳的师弟周伯通被东邪“黄药师”困在桃花岛的地洞里。为了打发时间,周伯通就用左手与右手打架,自娱

自乐。其武功诀窍在于要先“左手画圆、右手画方”,分心二用,保证可以同时使出两种武功,从而使得武力倍增。

这两部小说主人公功力大增的关键都是“自己跟自己决斗,试图倾尽全力击败对方”,而结果都是练成后,去跟高手过招时能轻松秒杀对手。用一句俗话来解释,叫“双拳难敌四手”。

## 让机器学会“左右互搏”

GAN 网络的原理本质上就是这两篇小说中主人公练功的人工智能或机器学习版本。

一个网络中有两个角色,修炼的过程中左手扮演攻方,即生成器(generator),试图生成和自然世界中拟完成任务足够相似的目标;右手扮演守方,即判别器(discriminator),试图把这个假的、生成的目标和真实目标区分开来。经过反复多次双手互搏,左手右手的功力都会倍增,从而达到“舍我其谁”的目标。

明白这个道理,就不难明白为什么 GAN 网络一出,就有独孤求败的感觉了。

正是因为 GAN 网络的机理是“双手互搏、一心两用”,所以,虽然

最初的应用场景是针对图像相关任务,但其机理是普适性的。只要能利用这个“诀窍”的地方,都能把自己的功力提升一个档次。不过,需要注意的是,GAN 只把双手互搏用来训练自己“双手”的功力。在多数实际应用中,它只用了自己千锤百炼出来的生成器,就得到了非常好的结果。

于是,自 2014 年 GAN 网络被 Ian J. Goodfellow 等人提出,至今其演绎出的各种版本就像攻城掠地般在各个领域蔓延。

在今年的人工智能顶级会议 IJCAI、机器学习顶级会议 ICML 和 NIPS、深度学习著名会议 ICLR 上,光读参会论文的标题,就能发现大

量 GAN 网络的工作痕迹。图像处理、计算机视觉、自然语言处理、语音识别、智能驾驶、安全监控……似乎 GAN 已经无所不能了。

张军平用几个比较火的人工智能应用场景举例——

在年龄估计中,GAN 能通过攻守互搏,基于给定的人脸图像实现对其人衰老或年轻时的样貌生成。

在多视角人脸识别、跨视角步态识别领域,采用类似的机理,实现对人脸角度和步态角度的自动旋转,从而有效提高了多视角、跨视角人脸和步态识别的精度。

在自动驾驶领域,需要对智能汽车进行复杂环境下的虚拟训练。此时,就可以利用 GAN 来实现与实

际交通场景分布一致的图像生成。具体来说,可以先给 GAN 输入一张随机噪声图像,通过其生成器来最小化与真实场景接近的图像,同时判别器最大化生成场景与真实场景间的差异。经过反复迭代的攻守博弈,从而获得与真实环境一致的交通场景。

“不光在应用领域上有了突破,互搏的方式也有了不小变化。既然可以双手互搏,那当然可以三手甚至更多手的互搏,也可以组团逐对互搏、成串互搏,诸如此类。还可以把圆和方换成其他东西或者所谓的函数或结构来互搏。但万变不离其宗,内在的机理是不变的。”张军平补充道。

## GAN 有什么“软肋”

“毫无疑问,这种可倍增‘功力’的技术入选 MIT 年度十大突破性技术是当之无愧的。但值得注意的是,这一技术仍然有较大需要改善的空间。”作为《IEEE Intelligent System》(智能系统)和《IEEE Transactions on Intelligent Transportation Systems》(智能交通系统)等著名期刊的编委,张军平对 GAN 的“软肋”了如指掌——

首先,对抗不好是容易出问题的。

比如那位 B 博士。国棋象棋的世界冠军在觉察到他的强烈焦躁与急切情绪后,就有意识地放慢下棋速度。结果诱发了 B 博士的精神分裂,使得他再次陷入狂乱的自我对弈之中,最终才幡然醒悟而告别了棋局。对抗网也是如此,稳定性一直是其存在的问题之一。尽管理论上 Wasserstein GAN (WGAN) 可以部分解决其收敛性问题,但实际的效果还没有达到尽人意

的地步。

其次,枪打出头鸟,既然这项技术这么牛了,过来挑战的也不少。据不完全报告,似乎已经有十五波“武林高手”对 GAN 网络的抗击打能力进行了挑战。极端情况,在图像中加一个像素就可能使 GAN 网络产生误判。

第三,GAN 网络也是深度网络的一种,在可解释性这条路上,仍然没

有找到非常明确的方向。

最后,双手互搏的基础还是手。而这一基础的结构并没有从近几十年来人工智能发展的框架中脱离出来。

因此,张军平提醒,期望以 GAN 之奇技,从弱人工智能冲击“能真正推理和解决问题,且有知觉、有自主意识的”的强人工智能还遥遥无期。

(据人民网)