

科普系列——强化学习及其在游戏测试中的应用

正如标题所示, 本篇文章主要介绍几年前 AlphaGo 和 OpenAI 用到的技术, 即强化学习, 以及它在游戏测试中的一些可能的应用。笔者也不太敢确定强化学习到底在游戏测试中做什么才能很好地落地, 因此本篇文章会从深度学习讲起, 逐步介绍强化学习, 以及展望一下强化学习在测试中能起到的作用

1. 深度学习

关于深度学习的概念, 这里就不做过多的介绍了。这里就举几个小例子, 来说明一下深度学习的作用归根结底到底是什么。为什么他这么神奇, 他是怎么做到的! 这里我们想象一下有一个二分类任务。

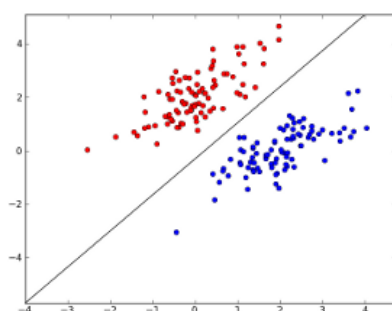


图 1 一个二分类任务示意图

在一个平面上, 有很多个点, 他们有两种颜色。如果要让人类去画一条直线, 分开红色和蓝色的点, 这显然非常的简单。我们随手一划就能划出一条完美曲线去区分两类点。在如上图所示中, 直线上方的是红色的点, 直线下方的是蓝色的点, 这非常的直观。但是如果要让机器去划分两类点, 这可就没这么容易了。当然我们现在已经发明了很多算法去做分类的任务了, 对于上图中的任务, 我们只需要对点的坐标进行建模, 然后计算机就能用数学方法, 快速的求出这一条直线了~。下面我们来看一个复杂点的例子

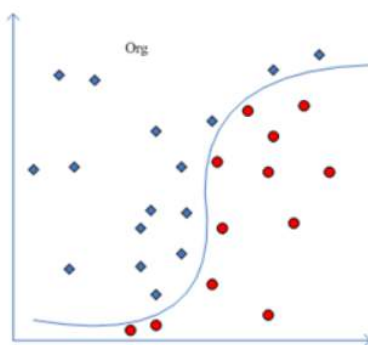


图 2 一个稍复杂的二分类任务示意图

对于上图中的这个分类就有点难度了! 但是依然难不倒人类, 我们只需要稍微拐一个弯, 就能完美的划出一条曲线去区分两类点! 但是机器的话可能就需要稍微复杂一点的方法了,

当然要计算整个曲线方程，现在人类还是发明了很多的数学算法的去计算这个曲线方程！比如逻辑斯特回归，支持向量机等等，这种任务依然难不倒计算机！

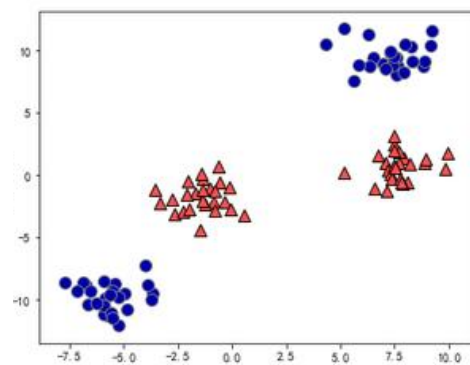


图 3 一个复杂的二分类任务示意图

接下来我们再看看一个更加复杂的例子，依然是一个二分类问题！对于上图这个任务我们人类可能要稍微思考一下，才能划出一个圆圈去区分这两类点了！但是计算机怎么办呢！计算机是如何区分的呢？这个时候我们就需要一个更加牛逼的函数了，也许这个函数在平面上，不能画出来，但是如果我们从高维的角度去看，我们依然是可以把两类点区分开来的。

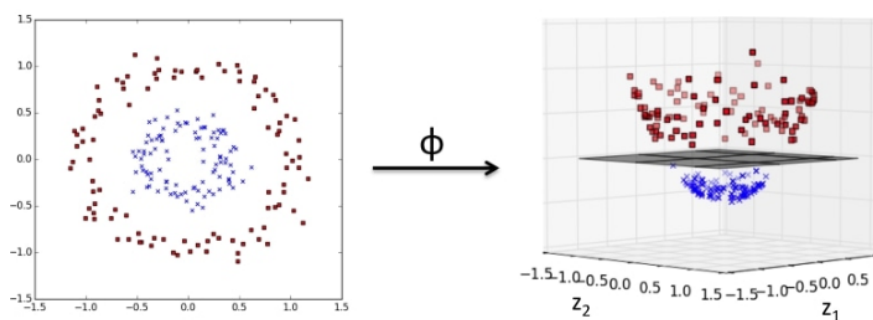


图 4 高维分类示意图

那么问题来了！怎么表示这个高维的函数！当类别数量越来越多咋办！也就意味着**我们需要的函数，越来越复杂**。人们已经无法用公示的形式去表达恐怖如斯的函数了！所以人们发明了一种算法，叫做**神经网络**，这个神经网络就能够拟合各种各样的函数~有理论表明，**一个三层的神经网络，就能够拟合大部分函数了**~关于神经网络的介绍，网上有很多解释，这里就不做过多的介绍了！

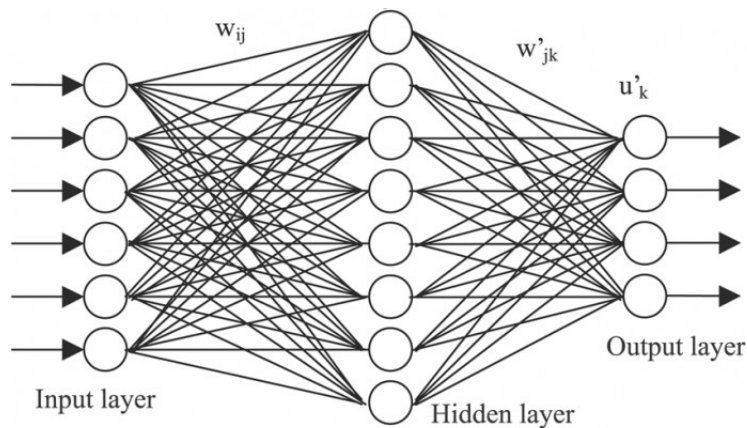


图 5 一个非常简单地神经网络示意图

既然有了神经网络，我们为什么还需要深度学习？因为简单地神经网络，还是不能够拟合一些超级复杂的函数，或者拟合的效果不是特别的好，这个时候，人们发现，只要把网络的层数加多，或者稍加设计一下，就会拟合得特别好！于是人们提出了深度学习的概念~所以我们可以简单的总结一下，**深度学习就是学习一个有很多层的神经网络。**

2. 强化学习

那么什么是强化学习呢？实际上在深度学习领域，根据**任务类型的不同**，人们把深度学习分成各个不同的方向，如**自然语言处理**，**图像处理**，**语音识别**等等。根据网络**学习方式的不同**，人们又把深度学习分为了三类，一个是**监督学习**，一个是**无监督学习**，还有一类就是**强化学习**。强化学习通过不断地与环境交互收集数据来训练自己。他可以用来解决行为决策问题。或许这样子说还是很抽象，我们现在举一个具体的例子。假设我们在玩阴阳师



图 6 阴阳师游戏画面

我们想要通关一个副本，我们就必须要放对技能。对于一个副本而言，我们人类如果熟悉阵容的话，非常快的就能学会什么时候该放什么技能！比如上图中，茨木应该对左边的雪女放大招，这样才能最快的通关副本。但是对于机器来说，要学会这个可不简单。我们不妨来看看我们人类是怎么决策的。我们人类在看到【现在是茨木在行动，并且对方三个还存活，并且雪女是最不肉的】所以我们要放大招打雪女。我们可以看成这是一个对应关系

茨木存活，雪女存在 —— 对雪女放大招

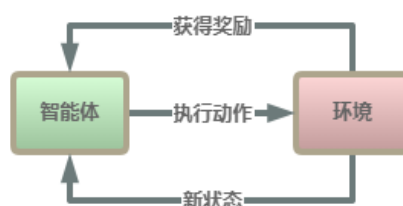
我们人类要学会这个对应关系，非常的简单，一下子就记住了。对于机器而言我们也可以做一个映射表，当遇到这个状态的时候，就对雪女放大招！但是游戏的状态不止这一个，可能是很多很多的。比如围棋就有数不尽的游戏状态，阴阳师也是。

表 1 阴阳师状态映射表

状态	最优决策
茨木存活，雪女存在	茨木对雪女放大招
茨木存在，大天狗残血	对大天狗放发招
茨木存在，剩下 Boss	对 Boss 放大招
茨木存在，没鬼火了	平 A 雪女
.....

但是阴阳师的每个式神的生命值和攻击力时刻都在变化，我们很难去列举完所有状态，也就是这个表格会十分的庞大！那么如何让机器学会这么复杂的映射关系呢？我们不妨把上面这个映射关系看成一个函数？函数的输入就是式神的生命值攻击力防御力等数值，输出就是最优的决策。这个时候我们是不是就能联想到一个东西？没错！就是深度学习！深度学习很适合用来解决复杂函数拟合问题！

那么这就是**强化学习**问题，我们对于每一个状态，需要让机器学到最优的决策，从而更好更快的通关副本。值的庆幸的是，这个问题已经得到了很多研究，并且已经有大量的成果了！比如最出名的 **AlphaGo**, **Open Five** 等，都是非常出名的玩游戏的人工智能。具体来说，我们已经有很多非常好的算法去解决这个网络训练问题，比如 **DQN**, **PPO**, **A3C** 等著名的算法。算法主体非常的简单，就是让机器不断地去试错，最终让自己记住每一个状态的最优决策。当然这个最优还是需要人为去定义的，何为最优，我们需要一个奖励函数，来计算机



器某一次行动获得了多少奖励。

图 7 强化学习示意图

如果一个智能体某次行为获得的奖励为负，那么智能体就会惩罚自己，让自己下次不再

输出这个行为，如果获得的奖励为正，那么智能体就会奖励自己，让自己下次更加可能的去输出这个行为。这就是强化学习的基本过程。由于本文是科普系列，就不多做强化学习算法的细节介绍了~有兴趣的同学可以自行上网查阅相关资料。

3. 能做什么？

现在我们已经知道，强化学习能够通过自我学习，不断的学习最优决策应该是什么。那么假设一个机器人学会了最优决策，那么这个最优行为决策，在测试领域又能怎么被应用呢？笔者在这里天马行空一下，简单地分析几个可能的应用场景。

3.1 性能测试

为什么说强化学习可以应用在性能测试呢？性能测试，无非就是就是在不同场景下，测试游戏的帧率，内存，网络等等的占用率，从而找到性能低下的地方，然后进行优化。

但是工具只能帮助我们监控，并不能帮助我们控制手机。在收集性能数据前，还需要我们人为的操作手机，去到某一个场景，然后再观察性能数据。那么这里就涉及到一个操作手机的操作。当然许多组都有很多 GM 指令，可以直接切换到某一个场景，但是如果要模拟真实玩家的情况，可能就需要人为操作了，比如 Moba 游戏，人为的模拟打团。但是如果我们有一个智能机器人，可以自动的帮我们走到某一个场景，或者自动的帮我们对战，那么就能解放很多人力。当然有的人会说，自动走自动打，我们可以自己写脚本，对于简单的样例来说，的确可以写脚本，但是对于一些复杂的逻辑可能就需要 AI 自动的去执行了。我们可以训练一个聪明的 AI，让他学会自动跑完整个场景，然后只需要调用 AI 来进行操作即可。比如在腾讯，这种测试方式已经被应用到了天天跑酷中！

3.2 生成残局，游戏内容生成

在游戏内容生成中，如果是闯关类游戏，可能需要策划进行关卡设计。当关卡数目较多时，这非常的花费人力，并且是一种重复的劳动。我们不妨利用 AI 去生成游戏关卡。强化学习算法非常适合用来生成具有挑战性的关卡~我们甚至还能使用 AI 自动的去验证策划设计的关卡或者生成的关卡的难度，大大降低了人力，和提高了游戏内容的丰富度。

3.3 更聪明的 NPC

在游戏中有许多的 NPC 和机器人，他们在游戏中都发挥着非常重要的作用。比如他们可以帮助完成游戏情节，增加游戏的挑战性，以及使游戏更加真实。传统上业界制作游戏的机器人和 NPC 基本上都是根据经验和规则，写成行为树，从而控制逻辑。这种方法的开发周期一般较长，成本比较大。最重要的是，获得的效果一般都不是特别的棒，而且当游戏比较复杂时，就不再适用了。这个时候我们不妨考虑使用强化学习去做，让 AI 自己学会一个比较真实的 NPC 的表现，或者机器人的表现。

3.4 平衡性测试

游戏数值的平衡性对于游戏的稳定运营至关重要，其对游戏造成的影响有时甚至超过一个严重的 bug。但平衡性设计，尤其是战斗系统的平衡性，往往很难考虑周全。作为 QA，如果我们利用强化学习算法，训练除了一个聪明的 AI，那么我们就能够利用这个聪明的 AI 做很多事情了！比如让 AI 自动的去玩一个副本，然后得出 AI 的胜率，从而评判一个副本的难

度。又或者让 AI 之间互相对打，得出双方的胜率，从而评判一个角色的技能或者数值是否合理。