014 | 精读AlphaGo Zero论文

2017-11-03 洪亮劼

AI技术内参 进入课程 >

讲述：初明明

时长 09:56 大小 4.56M

长期以来，利用人工智能技术挑战人类的一些富有智慧的项目，例如象棋、围棋、对话等等都被看做人工智能技术前进的重要参照。人工智能技术到底是不是能够带来“拟人”的思维和决策能力，在某种意义上成了衡量人工智能水平以及整个行业发展的试金石。

在这些项目中，围棋可以说是一个难度很大的项目，更是饱受关注。一方面，围棋的走棋可能性非常多，且棋局变化多端，这似乎给计算机算法带来了很大的困难。另一方面，围棋在一些国家和地区（比如中国、日本和韩国）不仅仅是一个智力竞技项目，更有一系列理念、人格等全方位的配套文化概念，使得人们对此产生怀疑，人工智能模型和算法是否能够真正学习到有价值的信息，而不仅仅是模拟人的行为。

2015 年，来自谷歌 DeepMind 团队的 AlphaGo 打败了职业二段樊麾，在这之后短短两年的时间里，该团队的人工智能技术迅猛发展，不仅以 4 比 1 击败韩国的李世石九段，更

是在今年战胜了即时世界排名第一的柯杰，可谓战胜了被誉为“人类智慧皇冠”的围棋项目。

前段时间，DeepMind 团队在《自然》杂志上发表了 AlphaGo 的最新研究成果，AlphaGo Zero，这项技术更是把人工智能在围棋上的应用推向了一个新高度，可以说是利用计算机算法把围棋的棋艺发展到了一个人类之前无法想象的阶段。

今天，我就来带你认真剖析一下这篇发表在《自然》杂志上的名为《不依靠人类知识掌握围棋》（Mastering the Game of Go without Human Knowledge）的论文 。标题中的不依靠人类知识当然有一点夸张，不过这也正是这篇论文的核心思想，那就是仅用少量甚至不用人类的数据来驱动人工智能算法。在之前的很多人工智能模型和算法来看，这是一个巨大的挑战。

作者群信息介绍

文章共有 17 位作者，都来自伦敦的谷歌 DeepMind 团队。AlphaGo 的第一篇论文也是发表在《自然》杂志，当时有 20 位作者，比较起来，这篇论文的作者数目减少了。另外，虽然两篇论文的主要作者都是三名（共同贡献），但是这三个人发生了一些变化。下面，我就介绍一下本文的三个主要作者。

第一作者大卫·希尔维（David Silver）目前在 DeepMind 领导强化学习（Reinforcement Learning）的多项研究。大卫的经历很传奇，早年曾在南非生活和学习，1997 年从剑桥大学毕业后，先到一家名为 Elixir Studios 的游戏公司工作了好几年。然后到加拿大的阿尔伯塔大学（University of Alberta）学习机器学习，特别是强化学习。他当时就开始尝试开发用计算机算法来进行围棋博弈。大卫 2013 年全职加入 DeepMind，之后迅速成了 DeepMind 在强化学习，特别是深度学习和强化学习结合领域的领军人物。

第二作者朱利安·施瑞特维泽（Julian Schrittwieser）是谷歌的一名工程师，他长期对围棋、人工智能感兴趣。值得注意的是，朱利安这次成为主要作者，而在之前的第一篇文章中还只是普通贡献者，可以推断在 AlphaGo Zero 这个版本里有相当多的工程优化。

第三作者卡伦·西蒙彦（Karen Simonyan）是 DeepMind 的一名科学家，长期从事计算机视觉和人工智能技术的研究。他来自 2014 年 DeepMind 收购的一家名为 Vision Factory 的人工智能公司。卡伦最近几年的论文都有高达几千的引用量。

论文的主要贡献

首先，**这篇论文的主要“卖点”就是较少利用、或者说没有利用传统意义上的数据驱动的棋局。**第一篇论文里的 AlphaGo 以及后面的一些版本，都是主要利用人类的棋局作为重要的训练数据，采用监督学习（Supervised Learning）和强化学习结合的形式。在 AlphaGo Zero 这个版本里，人类的棋局被彻底放弃，而完全依靠了强化学习，从完全随机（Random）的情况开始，“进化”出了具有人类经验的各种走法的围棋策略，并且达到了非常高的竞技水平。可以说这是本篇论文的核心贡献。

在核心的模型方面也有不少改进，比如一个很大的改进就是**把策略网络（Policy Network）和价值网络（Value Network）合并**，这样就能更加有效地用简单统一的深度模型来对这两个网络进行建模。另外，整个模型的输入特征也有变化，从深度模型提取特征外加人工挑选特征，到这篇文章提出的**完全依靠棋盘的图像信息来自动抓取特征**，可谓是减少人工干预的一个重要步骤。

文章的另一大看点是实验结果。作者们展示了新的 AlphaGo Zero 模型能够战胜之前很多版本的模型，最令人惊奇的可能莫过于 AlphaGo Zero 在“自学”的过程中，还“悟”到了很多人类在围棋学习过程中领悟的棋局招数。

论文的核心方法

AlphaGo Zero 模型的核心起源于一个简单的深度网络模型。这个深度网络的输入是棋盘当前位置的表达（Representation）以及过去的历史信息，输出一个走子的概率以及对应的价值。这个价值是用来描述当前棋手能够赢的概率。刚才我们已经说了，这个深度网络集合了策略网络和价值网络，形成了这么一个统一的评价整个棋盘的神经网络。在具体的网络架构方面，AlphaGo Zero 采用了计算机视觉领域最近流行的**残差架构**（ResNet），可以说也是这个方法的一个小创新。

有了这个基本的神经网络之后，作者们就需要和强化学习结合起来。具体来说，在每一个位置的时候，算法都会去执行一个**蒙特卡罗树搜索**（Monte Carlo Tree Search），对当前的神经网络模型输出的走子策略进行一个修正，或者可以认为是“加强”。这个蒙特卡罗树搜索的输出结果依然是走子的概率，但是这个概率往往比之前单从神经网络得到的要强。然后，更新神经网络的参数，使得参数尽可能地接近蒙特卡罗树搜索的结果。

那么，什么是蒙特卡罗树搜索？简单来说，就是我们从当前的棋盘情况出发，对整个棋盘产生的所有可能性进行有限制情况的搜索，也就是说，不是“穷举法”。大体说来，从某一个可能性走到下一个可能性主要是依靠下一个可能性发生的概率，以及通过神经网络来判断是否能赢的可能性。

整个算法最开始的时候是从随机的位置初始化，然后通过对神经网络的更新，以及每一个迭代通过利用蒙特卡罗树进行搜索，从而找到更加合适的神经网络模型的参数，整个算法非常简单明了。不管是结构上还是复杂度上都比之前的版本要简洁不少。文章反复强调公布的算法可以在单机上运行（基于 Google Cloud 的 4 TPU 机器），相比于最早的 AlphaGo 需要使用 176 个 GPU，也可以看到整个模型的进化效果。

方法的实验效果

AlphaGo Zero 的实验效果是惊人的。从模拟中看，大约 20 小时后，这个版本的模型就能够打败依靠数据的监督学习版本的 AlphaGo 了。而到了 40 小时后，这个版本已经可以打败挑战了李世石的 AlphaGo。也就是说，不依靠任何人类棋局，AlphaGo Zero 在不到 2 天的运算时间里，就能够达到顶级的人类水平。

除了可以打败之前的 AlphaGo 版本以外，这个版本相比于监督学习的版本，在大约 20 小时以后也可以更好地预测人类对战的走子。并且随着训练时间的推移，这种预测的准确性还在不断提升。

刚才我们也提到了，AlphaGo Zero 在自我训练的对战中，在不依靠人类数据的情况下，的确是发现了相当多的人类熟悉的对战套路。然而，有一些人类在围棋历史中较早发现的套路却没有或者较晚才在 AlphaGo Zero 的训练历史中习得。这打开了很多问题，比如发生这样情况的原因究竟是什么等等。

最后，作者们展示了 AlphaGo Zero 非常强大的实战能力，在和之前最强的 AlphaGo 版本，也就是 AlphaGo Master 的对战中，AlphaGo Zero 取得了 100 比 0 的绝对优势。而相同的 AlphaGo Master 与人对弈的成绩是 60 比 0。

小结

今天我为你讲了发表在《自然》杂志上的这篇关于 AlphaGo Zero 的论文，这篇文章介绍了一个简洁的围棋人工智能算法，结合深度学习和强化学习，不依靠人类的信息。

一起来回顾下要点：第一，关注这篇文章主要作者的信息，我们可以推断出文章的一些变化方向。第二，这篇文章有两大看点，一是很少或者几乎没有利用人类的棋局数据，二是得到了显著的实验结果。第三，文章提出的核心模型将策略网络和价值网络合并，与强化学习相结合。

最后，给你留一个思考题，有人说 AlphaGo Zero 并不是完全不依靠人类信息，比如围棋本身的规则就是很强的监督信息；再比如，不管每一步的走动如何，棋局最后是输是赢，依然是很强的信息。那么，AlphaGo Zero 到底是不是还是依赖了很强的数据呢？我们能不能把 AlphaGo Zero 看做是监督学习的产物呢？你怎么看？

欢迎你给我留言，和我一起讨论。

上一篇　013 | 精读2017年KDD最佳应用数据科学论文

下一篇　015 | 精读2017年EMNLP最佳长论文之一

精选留言 (2)

写留言

黄德平
2018-11-29

2

补充一点认识，zero中的神经网络使用卷积神经网络，这个是跟围棋本身的规则相关。具体来讲是，围棋每个地方都可以落子，而且局面上不同地方的计分规则是一样的。

展开

范深

2017-11-03

2

规则明确到无一例外，可以说是强监督了。只是以前的搜索方法还没到“评价”那步，就卡死了。我觉得Zero更多是视角和工程的创新，当然也很励志。

作者回复: 是的。