棋类运动在人工智能背景下的发展问题研究

——AI围棋对弈软件优化技术研究

王少锋

一、AI围棋历史概述

（一）AI围棋的一些主要里程碑

人工智能（Artificial Intelligence，AI）围棋的历史较长，1950年著名的信息论创始人克劳德·香农（Claude Shannon）发表了关于计算机博弈的论文，当时主要以计算机国际象棋为主要研究对象。第一个计算机围棋程序是A.L.佐布里斯特（A.L.Zobrist）于1968年开发。

20世纪80年代至90年代，出现了很多计算机围棋程序，采用的技术以极小极大树搜索、α—β剪枝、模式匹配、盘面评估函数、基于围棋领域知识的启发式搜索等传统AI技术为主。这一时期著名围棋软件有手谈、多面围棋（Many Faces of Go）、GnuGo等，但由于围棋本身的复杂性，以及缺乏有效、准确的方法来评估棋局盘面，因此采用传统AI技术的围棋软件棋力和人类职业棋手相比还较弱。

2006年，蒙特卡洛树搜索（Monte Carlo Tree Search，MCTS）技术开始应用于AI围棋，对AI围棋棋力有巨大的提升，MCTS对AI围棋软件是一个里程碑式的进步。自从以后，AI围棋软件棋力逐年提高，与职业棋手对弈时受让子数逐渐减少，并开始取胜。这一时期著名围棋软件有Zen，Crazy Stone等。2012年，Zen在

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

受让四子的情况下，战胜了日本武宫正树职业九段，2013年，Crazy Stone在受让四字的情况下，战胜了日本石田芳夫职业九段。在2015年之前，AI围棋的棋力已达到了业余高段的水平，但仍不能在非让子情况下战胜人类职业棋手。

2014年12月开始，2个独立研究团队开始将深度卷积神经网络（Deep Convolutional Neural Network，DCNN）应用于AI围棋。英国爱丁堡大学Clark和Storkey首先发表了论文，将8层结构的DCNN应用于围棋，用于预测人类棋手棋局盘面的下子点，准确率达到41%—44%，超过了以前其他方法的预测准确率[1]。10天后，加拿大多伦多大学和Google DeepMind公司AlphaGo团队也发表了论文，采用的神经网络相对稍大，是12层结构的DCNN，预测准确率可以达到55%[2]。

2015年10月，Google DeepMind的AlphaGo战胜了欧洲围棋冠军樊麾职业二段，2016年3月，AlphaGo的增强版战胜了韩国李世石职业九段，是AI围棋第一次在非让子情况下，战胜人类职业九段棋手，这是AI围棋又一个里程碑式的进步。这一时期著名AI围棋软件有AlphaGo系列软件，绝艺，星阵，ELF OpenGo，Leela Zero，KataGo等，AI围棋的棋力已完全超越人类棋手[3]。

（二）AI围棋待研究的问题

Google DeepMind发表AlphaGo系列论文后，已转向其他AI领域研究，使很多人误认为AI围棋问题已彻底解决。事实上，AlphaGo系列论文所解决的主要是计算机围棋中的搜索问题，所采用的方法是利用深度神经网络，使蒙特卡洛树搜索更高效，从而达到超越人类顶级职业棋手的水平。目前AI围棋仍遗留很多问题待解决。

1. 最佳一手问题

目前，AI围棋的棋力已远超人类棋手，但其所下子是否在任何情况下都是最佳一手，即是否是“围棋之神”下的子？根据现有围棋规则，如果禁止全局同形再现，理论上应该是能知道第一手棋下哪些位置是黑胜，下哪些位置是白胜。AlphaGo系列软件和目前已知的AI围棋软件都还不能解决该问题。

2. 对计算资源需求巨大

目前AI围棋对计算资源需求巨大，训练时间很长，而且训练时，往往不是训练一次，需要多次训练以优化参数。对于AI围棋，是否有更好算法，可以节省计算资源，并加快训练速度？

3. 难以精准分析一手棋的价值

围棋中的某手棋，有可能在当前看不出价值，但在几十手后才体现出这手棋

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的价值。目前AI围棋的网络架构还不能解决这类问题，包括AlphaGo在内的几乎所有已知AI围棋软件都是利用MCTS模拟分析一手棋的价值。MCTS是一种统计结果，在有限的计算时间和模拟步数内，得到的是一个近似值，是否存在更好的方法来分析一手棋的价值？

4. AI围棋的学习方式和人类棋手还有差异

AI围棋对超长大龙死活、复杂征子、双活等问题还不能很完美解决。里拉零（Leela Zero）是著名的AI围棋软件，多次在AI围棋大赛中获得好名次，棋力已远超人类职业棋手，但通过大量对局发现，LeeLa Zero遇到双活局面时，有时会莫名其妙自撞一气，导致赢棋变输。另外，征子对于人类棋手来说是个简单的概念，一般新手在刚开始学围棋时就很容易掌握征子战术。在AlphaGo Zero训练中，AlphaGo Zero可以很快掌握开局、死活、打劫、官子等围棋下棋技术，但对于征子则是在训练很长时间后才掌握。ELF OpenGo是由Facebook开发的另一款著名AI围棋软件，目的是为了复现AlphaZero算法。ELF OpenGo同样遇到了征子问题，在训练结束得到模型后，在判断局面是否征子有利时，仍然有一定的误判率[4]。对于征子问题的解决，目前还是需要依靠预先输入人类知识，如果完全从零开始通过强化学习，掌握征子仍是个困难的问题。

5. 存在棋力天花板

目前AI围棋自对弈训练所采用算法是深度强化学习，神经网络以卷积神经网络（或残差网络）为主，这种训练方法存在棋力天花板，即在模型参数规模确定的情况下，训练到最后，AI围棋的棋力难以再提高。是否存在更好的神经网络模型和算法，可以进一步提升AI围棋棋力？

6. AI围棋非技术性问题

除了AI围棋技术性问题外，还有其它非技术性问题待解决，如AI围棋对弈公平性、比赛规则等等。

上述待解决技术性问题中，最难或终极问题是第1个，即是否可以用AI实现“围棋之神”。

（三）国内外AI围棋研究

目前国内外有很多公司、学术机构及个人对AI围棋展开研究。国内的有腾讯AI实验室的绝艺，腾讯微信团队的凤凰围棋（PhoneixGo），北京深客科技有限公司的星阵，聂卫平围棋道场的棋精灵，南京栖霞棋院的清石围棋，清华大学的神算子，中国香港的MayaGo，中国台湾交通大学的CGI，国外的有韩国的韩

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

豆（HanDol），日本的GLOBIS　AQZ，美国的KataGo，比利时的Leela Zero，Facebook的ELF OpenGo，Google个人围棋爱好者发起研发的MiniGo，等等，虽然上述AI围棋有些已停止研发，如凤凰围棋、神算子等，但仍可以看出AI围棋是当前国内外的研究热点。在近几年的每次世界AI围棋大赛上，都有新的AI围棋软件出现。

二、AI围棋对弈软件技术分析

（一）AI围棋主要技术

研发一个AI围棋软件涉及技术较多，但主要是AI技术和蒙特卡洛树搜索这2个技术。除此之外，还会涉及其他相关的软件研发技术。

1. AI技术

目前计算机围棋中所用的AI技术以深度学习为主，所用的神经网络主要是卷积神经网络（Convolutional Neural Network，CNN）。随着深度学习技术的飞速发展，AI围棋所采用的网络也吸取了最新深度学习技术。比如，AlphaGo刚出现时，是采用一般架构的CNN，而且是采用多个神经网络，每个神经网络是单任务结构。而残差网络（Residual Network）提出后，由于其高性能的特点，近几年已成为主流的CNN，而多任务的单一神经网络又比多个单任务的神经网络性能好，因此AlphaGo之后的AI围棋，如AlphaGo Zero，AlphaZero，MuZero，ELF OpenGo，Leela Zero，KataGo等所采用的神经网络架构已改用多任务学习的残差网络。

2. 蒙特卡洛树搜索

自从2006年计算机围棋软件引入蒙特卡洛树搜索MCTS方法后，围棋软件棋力迅速提升，当前几乎所有的强AI围棋软件都已采用MCTS方法。

一般说来，采用MCTS方法时，模拟的次数越多，估计值会越准确，棋力也就越高。理论上讲，如果仅仅用MCTS搜索，最后也将渐近趋向最优下子点，但问题是这样模拟所需的计算量太大，在实际应用中不可能实现。而深度神经网络可以减少MCTS的搜索树宽度和深度，深度神经网络技术和MCTS这两种技术的结合，正是AlphaGo等棋力大幅提高的主要原因。

除上述2个主要技术外，要做好一个AI围棋软件，还需要相关的软件研发

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

技术。

3.多线程技术

软件棋力的高低和计算资源密切相关，多线程计算能充分利用计算资源。

4. 分布式计算技术

强AI围棋软件往往将后端引擎或网络模型部署在计算集群上，并通过远程过程调用返回当前棋局盘面下的最优下子点。AlphaGo，AlphaGo Zero等软件都有分布式版本，比相应的单机版本棋力要强。

5. 数据结构和算法

表示棋盘、棋局、走子、吃子等数据结构以及算法。比如表示棋盘，可以用二维数组，也可以用一维数组；棋盘上的一个点是黑子、白子还是未下子可以用一个整数表示，也可以用三个位（bit）表示。由于AI围棋中有大量和棋盘、走子、盘面判断等相关的运算，因此不同的数据结构对AI围棋软件的性能会有一定的影响。算法往往和数据结构紧密联系，若采用的数据结构不同，相应的算法也会不同。除了通用算法外，还有和人机对弈相关的特定算法，如用于判断两个棋局盘面是否一致的Zobrist哈希算法。

6. GTP协议

GTP（Go Text Protocol）是一个基于文本的传输协议，目的在于可以使得不同计算机围棋程序间可以自动进行对弈、计时、比赛结果的计算等。由于GTP的简单性以及易用性，目前AI围棋之间的对弈、AI围棋后端引擎和前端用户界面之间的通讯等，往往采用GTP协议进行。为了支持不同AI围棋软件之间直接对弈，以及方便用户的使用，AI围棋软件一般都支持GTP协议。

7. SGF文件格式规范

AI围棋软件除了对弈，还需要一些辅助功能，如支持棋谱保存、复盘研究等。目前SGF（Smart Game Format）文件格式已成为记录围棋棋谱的主要格式，AI围棋软件一般也都支持SGF棋谱的导入、导出等等。

8. 图形用户界面

AI围棋软件需要提供图形用户界面（Graphical User Interface，GUI）供用户方便使用。当前已有多款开源围棋GUI，如Sabaki，Lizzie，GoGui等，都支持GTP协议。开发者可专注于AI围棋后端引擎，不用再分散精力在GUI研发上。当然，如果认为开源软件的GUI不满足特定的需求，这时就需要自行研发GUI。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

9. 等级分计算机制

研发AI围棋软件是一个不断迭代过程，新版本软件是否强于以前版本，需要进行新旧版本之间的对弈并计算新版本的等级分并进行评估。计算等级分的方法较多，目前AI围棋中，计算等级分用得比较多的是BayesElo算法[5]。AlphaGo系列软件，每次推出新版本前，会和其它开源软件、人类棋手或先前发布的AlphaGo系列软件旧版本进行比较，其比较的一个方法就是采用BayesElo算法。

10. 围棋专业知识

虽然并不要求AI围棋研发人员必须懂围棋，但了解一些围棋规则和知识对研发AI围棋软件会有帮助，尤其是在测试阶段，围棋专业人士可以较快发现软件中缺陷或需优化的地方。

11. 软件工程技术

AI围棋软件的研发很大部分属于工程性质项目。如何保证软件运行稳定、性能高、易用、易维护、错误少，如何管理整个团队的研发过程等等，都需要软件工程技术。AlphaGo中所用的深度学习、MCTS等技术和算法并不是突然出现的，也不是Google DeepMind公司首先提出，其他公司和个人也一直尝试在AI围棋中应用这些技术。AlphaGo之所以最先开发出超过人类职业棋手的AI围棋，和Google DeepMind公司强大的工程研发能力有密切的关系。

12. 其他

以上是开发AI围棋软件技术的最小集合，如果要对软件进行商业化推广，还需要其他技术，如数据库、Web服务后端、前端、运维等等。

除了上述神经网络架构、算法、协议、规范等计算机软件层面技术外，计算机硬件资源，尤其是GPU计算资源也很重要。在AI围棋领域，硬件资源会极大影响架构、算法的权衡选择。

（二）AlphaGo系列软件技术概要

Google DeepMind公司发布的AlphaGo[6]，AlpahGo Zero[7]，AlphaZero[8]，MuZero[9]系列软件对AI围棋领域有巨大的影响，在此，我们先对AlphaGo系列软件的技术特点做一些简单介绍。

1. AlphaGo技术概要

AlphaGo的主要特色是基于监督学习和强化学习方法，训练深度卷积神经网络，找到了有效的棋局盘面评估函数，并采用了将神经网络评估值和MCTS相结合的新搜索算法。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

AlphaGo中使用的神经网络包括策略网络（policy network）和价值网络（value network）。首先根据人类棋手棋谱，基于监督学习方法训练策略网络，然后再通过强化学习优化提升策略网络，并同时训练价值网络。另外AlphaGo中还训练了一个较小模型的快速走子策略网络，用于MCTS模拟走子。

AlphaGo分单机版和分布式版本。单机版的AlphaGo首先采用人类棋手棋谱，用50个GPU训练了3周，得到监督学习策略网络，然后从监督学习策略网络到强化学习策略网络用50个GPU训练了1天，强化学习策略网络又用50个GPU训练了1周。而且训练神经网络模型，往往不是训练一次，需要多次训练以优化参数。分布式版本的AlphaGo用了280个GPU。因此，训练AlphaGo所需的计算资源非常巨大。

AlphaGo神经网络的输入使用了和围棋相关的一些知识，如征子、气、打吃等，因此AlphaGo不能应用于其他棋类。

2. AlphaGo Zero技术概要

2017年10月，Google DeepMind公司发布了AlphaGo Zero。AlphaGo Zero在AlphaGo的基础上做了很多改进，首先AlphaGo Zero将策略网络和价值网络合并为一个网络，而不是作为二个分离的网络分别训练；其次，AlphaGo Zero网络架构采用残差网络，而不是传统的卷积神经网络；第三，AlphaGo Zero训练时不再使用人类棋谱，而是采用强化学习自对弈的方式产生棋谱；第四，AlphaGo Zero在进行MCTS时，评估一个棋局盘面不再使用快速策略网络进行走子模拟，而是直接采用棋局盘面的价值预测值进行计算。

除了围棋规则外，AlphaGo Zero在训练时未使用其他围棋领域知识。和AlphaGo一样，AlphaGo Zero的训练也需要巨大的计算资源。

3. AlphaZero技术概要

2018年12月，Google DeepMind公司发布了AlphaZero。AlphaZero在AlphaGo Zero的基础上又进了一步，AlphaZero是一个通用算法，可应用于组合博弈中。组合博弈是指零和、完全信息、确定性的、按顺序下子的博弈游戏。组合博弈的范围很广，包括围棋、中国象棋、国际象棋、日本将棋等等都属于组合博弈。

AlphaZero在训练过程中对模型的更新方式和AlphaGo Zero不同。AlphaZero在迭代训练过程中，只维护一个神经网络模型，这个神经网络的参数被持续不断更新，自对弈棋谱总是由这个具有最新参数的模型生成。AlphaGo Zero在训练过程中会保留一系列的网络模型，自对弈棋谱由所有模型中最好的模型产生。在每个

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

迭代过程中，通过强化学习得到新模型，如果新模型对当前最好模型的胜率达到55%，则用新模型作为最好模型，以替换当前最好模型。

AlphaZero的主要特色在于其通用性，即同一个算法可以用于围棋、国际象棋、日本将棋或其他棋类。在AlphaZero出现之前，也有某些棋类软件，如国际象棋软件，其棋力超过人类顶级职业棋手，但软件中包含了大量领域知识，其采用的棋局盘面评估函数由人类专家做了仔细调优，因此只能用于该棋类。如果要用于其他棋类，需要重新开发。

4. MuZero技术概要

2019年11月，Google DeepMind首先在论文预印本平台arXiv.org发布，而后于2020年12月在《自然》（Nature）刊物上正式发布了MuZero。MuZero在AlphaZero的基础上更进了一步，在训练过程中，不再需要围棋的下棋规则，而是由机器学习得到一套下棋规则，通过自对弈掌握下棋规则。但训练得到的模型棋力比AlphaZero更强。MuZero相当于自己发明了一套“围棋”规则，这套规则可以用来下我们通常意义下的围棋，但MuZero对围棋的理解更深刻。

和AlphaZero一样，MuZero也是通用算法，可用于围棋、国际象棋、日本将棋以及雅达利（Atari）游戏。MuZero采用的残差网络比AlphaZero小，MuZero用了16个残差块（residual block），而作为对比的AlphaZero用了20个残差块，但MuZero性能却比AlphaZero还略好一些。

MuZero所采用的MCTS搜索算法和AlphaZero一样。但和AlphaZero不同的是，MuZero中的模型由三个部件组成，即将输入转换为内部状态的函数h，用于进行状态转移及计算奖赏的动态规划函数g，以及预测当前状态的策略值和估值的预测函数f。MuZero的创新之处是在训练时，将h，g，f这三个函数的参数同时进行训练，从而获得比AlphaZero性能更好的模型。

MuZero的计算资源需求也非常巨大，采用了谷歌云第三代TPU，总共1000个TPU用于自对弈，16个TPU用于模型训练。

三、AI围棋优化技术综述

（一）优化技术意义

通过分析AI围棋技术及AlphaGo系列软件所采用的技术，我们发现AI围棋的

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

关键技术是要找到一个准确且高效的评估函数，而深度神经网络可以起到评估函数的作用。AlphaGo系列软件整合了深度神经网络、MCTS，利用Google公司强大的计算资源，取得了成功。

在训练神经网络时，不管是采用自对弈方式产生棋谱，还是采用人类棋谱，都需要有巨大的计算资源，而这只有人力、资金雄厚的大公司才能提供，因此限制了AI围棋技术的推广和研发应用。

AI围棋技术优化可以降低对计算资源的需求，缩短训练时间，压缩模型大小，同时也可以提高棋力，是AI围棋众多问题中的基础性问题，若这个问题得以解决，可以吸引更多中小型公司和个人参与AI围棋研发中，而不仅仅限于拥有雄厚计算资源的大公司。

（二）优化技术分类

AI围棋的优化技术点较多，根据前面对AI围棋所涉及技术的讨论，我们将候选优化技术点分为神经网络模型、训练过程、MCTS三方面。下面从这三方面对当前可用的优化技术进行综合分析。

1. 神经网络模型优化

AI围棋的一个重要组成部分是深度神经网络，因此神经网络模型优化是重点。神经网络模型优化又分为以下几个方面。

（1）模型架构优化

深度学习是当前AI领域的研究热点，新的神经网络架构层出不穷，可以考虑从模型架构方面进行优化。

目前AI围棋主流的神经网络架构以残差网络为主，也有采用其他网络架构，如MobileNet等轻量级神经网络，也有采用Transformer网络架构的，如文献[10]采用和主流的卷积神经网络模型完全不同的GPT　2训练模型，但该方法还处于实验阶段，也未和MCTS集成，因此文献中仅讨论了采用GPT　2模型的AI围棋的可行性，但真实棋力并不清楚。由于Transformer网络架构具有很好的可扩展性，随着网络模型的不断增大，其性能会持续不断提升。2020年5月，OpenAI发布了GPT　3模型，GPT　3包含了1750亿个参数，参数量是GPT　2的100多倍，性能比GPT　2有极大提升。因此除了基于残差网络的AI围棋外，将来也有可能出现基于GPT　3的AI围棋。

对于卷积网络或残差网络，还可以考虑通过CNN插件进行优化。CNN插件具有类型众多、简单易用、对CNN性能有较大提升等特点，因此也是一个可行优化

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

技术。

（2）模型超参数优化

在网络架构确定之后，可以对神经网络模型的超参数进行优化设置。

AI围棋模型中的超参数较多，可分为以下几类。

第一类是和网络相关的超参数。如网络层数、每层的神经元个数、激活函数类型、权重初始化方法、L2正则化项的权重、各输出头部损失函数的权重，等等。比如AlphaGo系列软件中采用的激活函数主要是ReLU函数，但最新的研究表明，swish函数、h_swish函数等作为激活函数效果更好，因此可以用swish或h_swish函数代替ReLU函数进行优化。对于权重初始化方法，AlphaGo系列软件主要是采用随机初始化方法，但最新研究表明，kaiming初始化比一般随机初始化效果要好。文献[11]从AlphaZero中选取了12个超参数，在6 6奥赛罗棋（Othello）上进行实验调优，对各种超参数组合做了对比分析，对每个超参数用不同的3个值进行对比，总共运行了36次，得出这些超参数取哪些值性能更好。但这种做法很耗费时间，效率不是很高，更好的方法是采用神经网络架构搜索（Neural architecture search，NAS）技术自动搜索超参数的最优值。NAS是一种自动机器学习技术，目的是自动确定神经网络架构及合适的神经网络超参数。NAS也是当前深度学习领域的一个热门研究方向。

第二类是和训练方法相关的超参数。训练数据的批大小、学习率大小、强化学习的迭代次数、每次迭代参与训练的棋谱数量、自对弈时开局选择下子点的温度值（温度值越小，下子点越确定）、每步走子的MCTS模拟次数，等等。需要说明的是，和训练方法相关的超参数是和具体AI围棋相关，比如有些超参数，在AlphaGo Zero中存在，但在AlphaZero已去掉，如新网络替换旧网络要求达到的胜率值。

第三类是和MCTS相关的超参数。MCTS模拟时用于权衡选择未知待搜索节点和已搜索节点的常量、搜索的宽度和深度、并发搜索时虚拟损失值（virtual loss），等等。

（3）模型压缩

在模型架构、超参数等方面优化完之后，可以考虑进行模型压缩。模型压缩是指对已经训练好的模型进行精简，得到一个轻量且性能和压缩前模型性能差别不大的网络。压缩后的模型具有更少的参数，可以有效降低计算和内存开销，有助于模型在生产环境中部署。模型压缩是深度学习领域的一个热门研究方向，主

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

要技术包括紧凑模型架构设计、参数裁剪、参数量化、哈夫曼（Huffman）编码、减少CNN模型卷积核、知识蒸馏，等等。

AI围棋在进行MCTS模拟时，需要依赖神经网络的计算，小神经网络模型可以加快模型的计算速度，因此，模型压缩的另一个好处是可以增加MCTS模拟次数。一般地，神经网络模型越大，包含的参数越多，其预测准确率会越高，棋力也越强，但计算的开销也越大；另一方面，模型小，计算的开销小，在进行MCTS模拟时，可以在相同时间内，模拟更多次数，也会使得棋力更强。因此需要在神经网络模型大小和MCTS模拟次数之间有个平衡。

另外模型压缩后，还可以方便部署在移动设备上。大模型由于计算量大，能耗高，不适合部署在移动设备上。

一般说来，模型压缩后需要重新训练，但重新训练不需要在原来全部数据集上进行，可以采样部分数据进行训练。

2. 训练过程优化

从AlphaGo的监督学习，到AlphaGo Zero，AlphaZero，MuZero的强化学习，都非常耗费计算资源。一般只有拥有巨大计算资源的大公司才能提供，或者采用类似Leela Zero的方式，由全世界志愿者提供算力。

混合不同模型的权重，得到新模型可以提高棋力。例如，可以在同一数据集上训练多个模型，然后取这些模型的预测平均值。在机器学习领域，这是一种简单，但又非常有效的优化技术，几乎可以应用于所有的机器学习算法中。在AI围棋领域，已有此类相关的实践，如将Leela Zero不同版本模型的权重混合后得到新模型，其棋力超过了混合前各个版本模型的棋力。

3. MCTS方法优化

除了神经网络外，AI围棋的另一个重要组成部分是MCTS。MCTS的目的是提供对棋局盘面更准确的评估。MCTS实际上是一个状态空间搜索问题，一般分为四个阶段，即选择（selection）、扩展（expansion）、模拟（simulation）和回溯（backpropagation）四个阶段。

（1）选择阶段

从根节点开始，按一定策略，沿着树往下找子节点，直到找到一个当前优先级最高的可扩展节点。可扩展节点是指该节点不是最终状态节点，且有未被访问的子节点。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

（2）扩展阶段

对选择阶段确定的可扩展节点，从其未被访问的子节点集合中，选定一个或多个子节点，将其从未被访问的子节点集中移除，并添加到树中。

（3）模拟阶段

从扩展阶段所添加的节点开始，按指定策略模拟到达最终状态，得到一个模拟得分，比如棋局的胜负结果。这里模拟时的指定策略可以是随机的，也可以包含领域知识，比如优先模拟可打吃对方子的子节点。

（4）回溯阶段

根据模拟阶段的得分，更新当前子节点及其所有祖先节点的模拟次数、得分值等统计信息。

MCTS方法的每个阶段都可以进行优化。优化的思路是减少每步搜索时的宽度和深度，MCTS性能在很大程度上是由模拟时选择子节点的策略决定的。目前由于LeelaZero，KataGo等一些强AI围棋软件已经开源，有很多基于这些开源软件进行改进优化的AI围棋，其中较多的是从MCTS方法方面进行优化。

除了上述神经网络模型、训练过程、MCTS优化技术外，另外还有一些通用的优化技术，这些通用优化技术不仅适用于AI围棋，对所有的神经网络模型都有提升作用，比如根据特定硬件平台重新编译Tensorflow源代码以提升性能、采用多GPU实现、软件代码优化、硬件性能改进等等。本课题主要关注和AI围棋相关的优化技术，不对这些通用优化技术详细展开讨论。

（三）目前AI围棋中使用的优化技术

目前棋力较强的开源AI围棋软件有ELF OpenGo，Leela Zero，KataGo等。由于ELF OpenGo的目标是复现AlphaZero的实现方法，不是对AlphaZero进行优化，因此我们重点对Leela Zero和KataGo中的优化技术进行分析。

1. Leela Zero用的优化技术

Leela Zero是较早出现的开源强AI围棋软件。Leela Zero的主要创新之处是其训练方式，由全世界志愿者贡献空闲计算资源，产生训练数据，然后训练神经网络模型。Leela Zero棋力较强，多次在世界AI围棋大赛中获得前四名。其所采用的优化技术主要有以下这些。

（1）网络模型逐步扩大

Leela Zero在训练网络模型时是从小到大逐步训练，训练大网络时是基于小网络的训练结果再扩展训练得到，这样不用从零开始重新训练大网络，节省了大量

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

训练时间。

（2）输入通道增加

Leela Zero的输入采用了18个通道（channel），比AlphaGo Zero多了一个输入通道，解决了AlphaGo Zero中黑可以看到边界，而白不能看到边界的缺陷。

2. KataGo用的优化技术

KataGo是目前用得较多的开源AI围棋软件[12]，有很多不同于其它AI围棋的特色，比如支持多种围棋规则、支持让子棋、支持不同大小的棋盘，等等，其中令人印象深刻的是KataGo采用了多种优化改进技术，大幅度减少了训练时间。KataGo所使用的优化技术分二部分：不依赖围棋领域知识的优化技术和基于围棋领域知识的优化技术。

（1）不依赖围棋领域知识的优化技术

不依赖围棋领域知识的优化技术包括以下几种技术。

第一个技术是MCTS模拟（playout）次数最大值随机化（playout cap randomization）。Playout是指MCTS模拟时探索的过程，在MCTS选择阶段，每展开一个节点就是一次playout。显然模拟每步棋时playout次数越多，估计值就会越准确，但所需的计算资源也越多。AlphaGo Zero的playout次数设置为800，KataGo对此进行了优化，即不要求每步棋都进行同样次数的playout，而是每步随机进行不同次数的playout。比如一局棋有300步，其中随机50步进行最多600次playout，剩余的250步只进行最多100次playout，这样可以加快训练棋谱数据的生成。最后进行了600次playout的50步作为强化学习训练数据，其余进行了100次playout的250步不作为训练数据。

第二个技术是策略目标裁剪（policy target pruning）。和AlphaZero一样，KataGo采用各子节点的模拟次数分布作为策略网络训练目标，KataGo的改进是在MCTS模拟时强制各子节点至少进行最少模拟次数，同时模拟结束后将明显不好的子节点从训练目标中裁剪掉。

第三个技术是全局池化（global pooling）。在围棋中，存在非局部战术，比如打劫，而残差网络的卷积核关注的是局部信息。KataGo在残差网络中加入全局池化技术，以方便从输入数据中抽取全局信息。

第四个技术是附加策略目标（auxiliary policy targets）。KataGo除了考虑当前局面的下子策略外，还考虑当前局面下子后，对手的下子策略。该技术在AlphaGo之前已在AI围棋领域使用，但主要是在监督学习中使用，KataGo将该技

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

术移植到AI围棋的强化学习中。

（2）基于围棋领域知识的优化技术

基于围棋领域知识的优化技术包括以下几种技术。

第一个技术是分解训练目标。将价值网络的训练目标分解为地域归属、子数，等等，该技术不是KataGo首创，但KataGo进行了细化。和AlphaZero相比，采用这个技术能加快学习速度，明显改善强化学习效率。直观想象，如果只给出棋局盘面的价值，网络只能“猜测”棋局中哪个区域影响了最后的价值，而给出地域归属值后，网络能直接“知道”哪个区域导致价值网络的输出偏差，这样网络能以更少训练样本得到相同或更好的结果。

第二个技术是和围棋特征相关的输入。KataGo采用了气、征子等这些围棋领域知识的输入。AlphaZero没有使用围棋领域知识，是为了证明在没有围棋领域知识的情况下，通过AI强化学习技术也可以达到超过人类的水平，但并不排斥在AI围棋强化学习时使用领域知识。

KataGo这些优化技术效果明显。作为对比，AlphaZero用5000个TPU训练了数天，约合41个TPU年；ELF OpenGo用2000个V100 GPU训练了13　14天，约合74个GPU年；KataGo的训练采用最多28个（平均26　27个）V100 GPU，训练了19天，约合1.4个GPU年，KataGo大约减少了50倍学习时间。

KataGo主要是对神经网络模型架构及训练过程进行优化，但在模型压缩、MCTS这二方面考虑得不多。

3. 其他优化技术

除了Leela Zero，KataGo所采用的上述优化技术外，还有其他优化技术，如文献[13]采用基于群体训练（Population Based Training，PBT）方法加速改进AlphaZero的训练。由于调试超参数非常耗时，每个参数的调整都需要重新进行自对弈，而PBT方法可以只需运行一次。PBT是从训练方法上进行优化，在进行自对弈时，同时有16个智能体（agent）供选择，而AlphaZero只用了1个agent进行自对弈。PBT通过同时训练多个网络，对超参数进行调优，达到改进性能的目的。

文献[14]是对MCTS方法进行优化，采用“刚好赢”策略（Exact Win Strategy），利用子树的信息（胜、负、平、未知）以裁剪没必要的走子探索。

文献[15]提出双重MCTS算法，产生两个搜索树，即一个小搜索树和一个大搜索树。小搜索树可以模拟更多的次数，产生的结果可以作为节点的优先级排序，然后这些优先级可以用在大搜索树中，再通过大搜索树模拟得到最终结果。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

（四）优化技术选择准则

由于AI围棋涉及的技术点较多，可供选择的候选优化技术也较多。本课题选取其中的1、2个技术点作为研究目标。选取的标准是：

首先，优先考虑有助于提高棋力、提高网络输出预测准确率、加快训练速度的优化技术。

其次，优先考虑和AI围棋领域相关的技术。比如神经元丢弃（dropout）技术，在某些AI领域用得比较多，可以解决过拟合问题，提高模型的预测准确率，但在AI围棋领域，由于训练数据足够，过拟合不是主要问题，且采用dropout技术会大幅度增加训练时间，因此在AI围棋中用得不多。本课题暂不考虑这类优化技术。

第三，不选取其他人已实验过、已证明可行的优化技术，如MCTS优化技术，即不重复其他人工作。

基于以上的选取准则，本课题研究采用知识蒸馏技术作为研究对象。

四、知识蒸馏优化技术

（一）知识蒸馏优化技术概述

知识蒸馏概念最先由布奇卢阿（Bucilua）等人于2006年提出[16]，其想法是将多个相关模型集成到单一模型中，从而达到模型压缩、易于部署的目的，但当时深度学习还不流行，知识蒸馏方法并没有引起人们的重视。直到2015年，辛顿（Hinton）在文献[17]中对知识蒸馏概念做了进一步拓展，知识蒸馏才开始引起人们的注意，并在近几年成为研究热门。

知识蒸馏是一种神经网络压缩和加速方法，可以有效提高神经网络的性能。具体来说，知识蒸馏是指用一个较小模型去拟合一个较大模型，从而让小模型学到与大模型相似的函数映射。采用知识蒸馏方法比直接在原始数据上进行训练，得到的模型性能更好。和其它模型压缩方法不同的是，知识蒸馏对小模型和大模型，在网络架构上并没有特殊要求。

一般地，对于残差网络，网络越深，模型的预测能力越强，基于此模型的AI围棋棋力也越强；同样深度的网络，卷积核越多，棋力也越强。但网络越大，对部署环境的要求也越高，且在应用时，要求的计算量也越大。所以较好的方法是

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

先用大网络训练，训练完成之后，通过知识蒸馏方法，将知识迁移到小网络中，从而达到压缩模型的目的。

在知识蒸馏中，往往将待蒸馏的大模型称作教师模型，将蒸馏得到的小模型称作学生模型。为了更好理解知识蒸馏技术，我们以人类社会中学生跟从教师学习的模式作类比。

首先教师用大量时间和精力学习掌握到知识，然后再用较短的时间将知识教给学生。由于学习时间较短，大部分学生学到的知识会比教师欠缺，但少部分学生，因天赋、学习方法、课后复习、补习、自学等原因，会青出于蓝而胜于蓝，在知识掌握上超过教师。这里学习时间相当于模型训练时间，天赋相当于精巧的模型架构及超参数，学习方法相当于模型训练方法，课后复习相当于更多的模型训练时间，补习相当于更多教师模型的输入，自学相当于强化学习。

学生之间还可以组成兴趣小组，任何一个学生可以充当教师角色教其他学生，共同学习提高，这类似于知识蒸馏中多个模型之间的互学习。

教师教学生的同时，自己也得到了提高，即教学相长，这类似于知识蒸馏中教师模型和学生模型共同优化。

知识蒸馏的算法较多，如对抗式蒸馏（Adversarial Distillation）算法受生成式对抗网络（Generative Adversarial Networks）启发而提出，类似于教师出题、学生解题的过程；多教师蒸馏（Multi Teacher Distillation）算法，不同的教师模型可以给学生模型提供不同的知识，在训练过程中，多个教师模型可以同时训练一个学生模型。这也类似于人类社会中，一个学生可以跟从多个老师学习；跨模态蒸馏（Cross Modal Distillation）算法，教师模型和学生模型的输入数据形态属于不同类别，即教师模型和学生模型可以是处理不同类型的任务，比如教师模型学习到的是关于图像识别的知识，而学生模型是对文本的识别；无数据蒸馏（Data Free Distillation）算法，教师模型采用人工合成数据的方法来训练学生模型。虽然无数据蒸馏算法有很大的应用前景，但如何生成高质量人工数据仍是一个有很大挑战性的问题；量化蒸馏（Quantized Distillation）算法，即学生模型采用量化网络。网络量化（network quantization）是指将高精度（比如32位比特浮点数）网络转换为低精度（比如2位或8位比特整数）网络，从而降低神经网络的计算复杂度。其它的还有基于注意力蒸馏（Attention Based Distillation）算法，基于图蒸馏（Graph Based Distillation）算法，终身蒸馏（Lifelong Distillation）算法，基于NAS蒸馏（NAS Based Distillation）算法，等等[18]。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

（二）知识蒸馏方法特点

对于神经网络中的知识蒸馏方法，可以用石油炼油做类比。在石油炼油过程中，根据石油中各混合物的沸点不同，利用蒸馏技术分别提炼出不同的产品。类似地，知识蒸馏是将教师模型中的“知识”蒸馏出来，并迁移到学生模型中。在知识蒸馏中，同样有温度设置问题。提炼不同神经网络模型中知识的最佳温度值，一般不会相同，知识蒸馏中的温度设置依赖于实践经验值。

和其他模型压缩及优化方法相比，知识蒸馏方法有自己的特点：

1. 适用性

知识蒸馏技术具有很好的适用性。在用教师模型训练学生模型时，学生模型既可以是参数量少很多的小模型，也可以是和教师模型参数量差不多的大模型，而且知识蒸馏方法并不限定采用哪种网络架构，可以将残差网络替换为其它轻量级网络，如MobileNet，EfficientNet，Xception等等，仍然可使用知识蒸馏。

2. 训练数据量少

知识蒸馏训练所需要的数据可以少于采用非知识蒸馏训练所需要的数据。例如，训练“尧弈”教师模型（神经网络为12个残差块，每层192个卷积核）时，所用的数据量为2700多万个棋局盘面，而在训练“尧弈”学生模型（神经网络为4个残差块，每层128个卷积核）时，只使用了约420万个棋局盘面。

3. 兼容性

知识蒸馏方法具有较好的兼容性。使用知识蒸馏方法后，并不排斥进一步使用其它优化技术。知识蒸馏方法可以和其他神经网络压缩技术，如网络剪枝、量化、权重聚集等一起使用，也可以和其他神经网络模型优先技术、训练过程优化技术、MCTS优化技术一起使用。

（三）知识蒸馏在AI围棋中的应用

知识蒸馏已在AI的其他领域，如人脸识别、图像/视频分割、目标检测、语音识别、自然语言处理等领域得到应用，但在AI围棋领域用得还不多。

AlphaGo系列软件并没有使用知识蒸馏方法，由于Google DeepMind公司拥有巨大的计算资源，可以专注于模型预测准确率的提高，模型压缩优化不是重点考虑的问题。而对于中小公司来说，不仅要考虑模型的预测准确率，还要考虑模型的计算开销。

本课题对AI围棋技术优化思路是：以AI围棋软件“尧弈”技术框架为基础，先训练一个大的AI围棋神经网络模型作为教师模型，然后通过知识蒸馏方法得到

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

一个小的学生模型。学生模型将和具有相同架构、网络层数、卷积核数量、训练数据和训练时间的非知识蒸馏小模型对比，比较模型的预测准确率及棋力强弱，以验证知识蒸馏技术在AI围棋领域应用的可行性及有效性。

（四）知识蒸馏在“尧弈”研发中的应用

1. 尧弈技术框架

尧弈是完全自主研发的AI围棋软件，包括核心引擎和客户端二部分，支持GTP协议和人机GUI对弈。尧弈基于AlphaGo系列论文所提出的原理，对其中的算法及实现技术进行了优化，以下是一些主要改进地方。

（1）训练数据改进

AlphaGo监督学习策略网络采用的训练数据是围棋对弈网站KGS上高段棋手的棋谱，AlphaGo Zero，AlpahZero，MuZero是采用自对弈棋谱数据。因KGS上有业余棋手数据，且其中包括了约35.4%的让子棋棋谱，棋谱质量并不是很高，而自对弈棋谱的生成需要巨大的计算资源。尧弈没有采用KGS棋谱数据，而是采用中日韩职业高段活跃棋手的棋谱，棋谱质量要比KGS棋谱高，而准备数据的时间要远少于AlphaGo Zero，AlpahZero和 MuZero。

（2）神经网络架构改进

本课题中尧弈采用了三个输出头的残差网络，模型输出包括二部分，即策略值输出和价值输出，其中策略值输出又分为下一步走子策略和下下一步走子策略，即网络总共有三个输出。尧弈模型中的激活函数采用了更有效的h_swish函数，AlphaGo系列中采用的是ReLU激活函数。在尧弈的前期研发过程中，我们发现采用h_swish激活函数得到的神经网络模型，比ReLU在预测准确率上有明显提升。尧弈模型的权重初始化函数采用kaiming初始化，AlphaGo系列中采用的是随机初始化。已有研究表明，在神经网络训练过程中，采用kaiming初始化比随机初始化效果要好。

（3）盘面价值判断方法改进

在AlphaGo系列中，盘面价值是根据棋局的结果确定的，这样，一局棋的所有盘面，其价值都是相等的。但在实际对局中，一局棋往往会出现反复，有时是黑优势，有时是白优势。尧弈在判断盘面价值时，不是简单根据棋局结果，而是根据当前盘面的形势判断确定，这样可以更精确判断一个棋谱所有盘面的价值。

（4）输入数据特征改进

为了加快训练过程，尧弈输入数据中采用了围棋领域知识。本课题研究训练

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的尧弈模型，输入通道和AlphaGo一样，共有49个通道，但每个通道的含义并不一定完全一致，例如，对于征子，尧弈中将其扩展为对所有气为2的棋块的判断；对禁着点的含义，尧弈是根据盘面状态和围棋规则确定。AlphaGo对输入通道的选取，是根据先验知识确定，对这些输入通道的描述也较简单。模型本身并不排斥选取不同的特征输入。本课题之所以也选取49个通道，是为了便于和AlphaGo的结果进行分析比较。

（5）其他改进

尧弈在MCTS搜索、训练方法、底层数据结构、分布式计算、开发框架等方面也采用了很多自研技术，由于和本课题研究内容重点不同，限于篇幅在此不展开论述。

2. 尧弈知识蒸馏算法

本课题中，尧弈所采用的知识蒸馏算法如图1所示：

首先用全部数据量训练“尧弈”大模型作为教师模型T。

第二步是从全部训练数据D中随机抽取部分数据d，用数据集d训练“尧弈”基准模型B。

第三步是设置温度值w，比如设置温度为2，用教师模型T对数据集d进行预测，获得策略网络输出，记为r，作为软标签（soft labels）训练数据集。

第四步是训练知识蒸馏学生模型，即构造具有和基准模型B相同架构、网络层数、卷积核数量的模型S，将数据集d和软标签训练数据集r一起输入到模型S进行训练，得到温度为w的知识蒸馏学生模型S。

第五步是比较模型B和S的性能，主要是比较B和S在测试数据集上策略值的预测准确率，及模型B和S之间直接对弈的胜负结果。

最后根据B和S的比较结果，判断是否需要继续实验。如果需要，则修改温度值，并重复上述第三至第五步过程。

需要说明的是，本课题研究中暂仅考虑策略网络输出部分的知识蒸馏，因策略网络输出属于分类问题，其输出端的逻辑回归输出适合于进行蒸馏，而价值网络部分仅包含一个输出值，如果要进行知识蒸馏，其采用的蒸馏方法会不同。

下面我们以2019年1月15日，中国柯洁九段执黑与韩国申真谞九段在第4届百灵杯决赛三番棋第1局为例，具体说明尧弈中知识蒸馏的训练过程。

任意选取棋谱中的一个盘面，例如，在下完第100手（图2中带×符号的白子）之后，盘面如图2所示：

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

这时轮到黑走子，实战中黑第101手将下在A位，然后白第102手下在B位。该盘面作为输入，黑在A走子、白在B走子、盘面胜负结果作为输出，将作神经网络的一个训练数据。尧弈训练模型时使用的是2700多万个类似这样的训练数据（具体数据量参看本课题报告第五部分）。

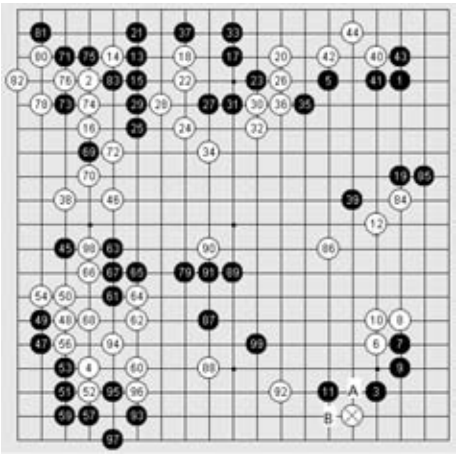

图2　柯洁–申真谘第4届百灵杯决赛第1局前第100手盘面

用这2700多万的数据量训练尧弈后，我们得到一个神经网络模型。然后用这个神经网络模型，预测在图2盘面情况下，黑将下在哪个位置。预测结果如图3所示：

在图3中，我们看到，黑下在A位的概率是0.934，除此之外，下在其它位置的概率分别是0.061，0.002，0.001，……，等（另外还有三个选点的概率分别是0.000432，0.000364，0.000264，四舍五入后在图3中显示为 .000）。

其中概率值最高的点和其它点的概率值相差很大，表示黑有很大可能下在A位。实战中，确实黑也是下在A位。但除了A位外，下在其它位置的预测概率，其实也是很有用的知识。知识蒸馏的目的就是将教师模型中的这类知识传授给学生模型，从而提高学生模型的预测准确率。

需要说明的是，在原始棋谱中，除了下一手是下在A位的信息外，并没有下在其他位置的信息。也就是说，下在其它位置的概率是尧弈教师模型在训练过程中获得的，是根据2700多万个棋局盘面的数据，训练后抽取出来的知识，不同于原始棋谱中包含的知识。

尧弈模型对下子位置的预测值是在神经网络的最后一层，通过软最大（softmax）函数输出得到。Softmax函数的作用是对预测值进行概率规范化，其定义是：

$$q_i = \frac{exp(z_i/T)}{\sum_j exp(z_j/T)}$$

其中zi是原始预测值，qi是规范化后的概率值，T是温度参数，上述公式的含

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

义是将逻辑回归的第i类的输出值zi规范化为概率值qi。公式中T通常默认设置为1，图3中的预测值即是温度T为1时的结果。我们发现，下A点的概率是0.934，下其它点的概率分别是0.061，0.002，0.001，……，等，即下A点的概率远远超过下其它点的概率。在训练学生模型时，这种差距过大的概率值，对训练模型虽然有用，但用处不是很大，如果提高温度T，则各个下子点的预测值将变得平缓，图4至图6分别是温度为2，3和5时的预测值，可以看到，随着温度的提高，各个点预测值之间差距变小，这样的预测值对训练学生模型更有用。

图4中各个下子点的预测值分别是0.572，0.147，0.028，0.016，0.012，0.011，0.010。和图3中温度为1时的预测值相比，各个点的概率值大小顺序保持一致，但相互间的差距变小了。

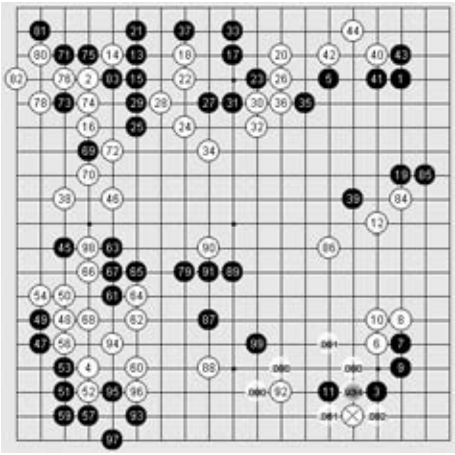

图3　尧弈对黑101手的预测值

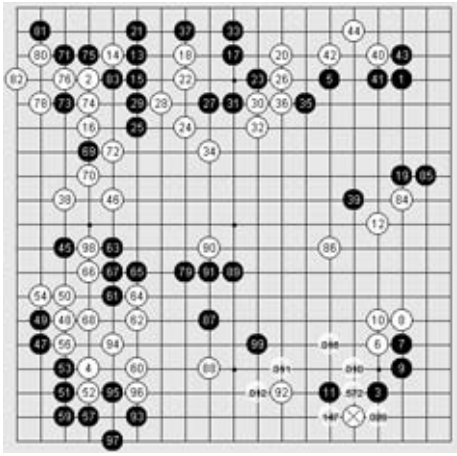

图4　温度为2时的预测值

图5中各个下子点的预测值分别是0.199，0.081，0.027，0.019，0.015，0.015，0.013。和图4中温度为2时的预测值相比，各个点的下子概率值大小顺序仍然保持一致，但相互间的差距已经不是很明显了。

图6中各个下子点的预测值分别是0.044，0.026，0.013，0.011，0.010，0.009，0.009。和图5中温度为3时的预测值相比，各个点的下子概率值大小顺序仍然保持一致，但相互间差距进一步缩小。

可以发现，随着温度值的升高，原来概率值差距很明显的下子点，差距变得越来越不明显，第一概率值和第二概率值在温度为2时，分别是0.572和0.147；温度为3时，分别是0.199和0.081；到温度为5时，则分别是0.044和0.026。其它点的预测概率值，也有类似的趋势。

在进行知识蒸馏时，各个点的预测概率值差距过大或过小，都不利于学生模型

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

的训练，需要设定一个最优的温度值T，而最优温度值往往是根据实践经验或实验获得，并没有一个标准方法。下面我们用实验方法确定一个较理想的温度值。

五.实验环境及结果评估

（一）实验环境

本课题实验的软硬件环境是：操作系统为Windows 10 Home版本；内存大小为48G；CPU为酷睿i7 7700K；GPU为NVIDIA 1080Ti；深度学习框架为TensorFlow 1.14.0和Keras 2.2.4。

（二）网络架构

教师模型是输入为49个通道，输出为3头部的残差网络，神经网络包含12个残差块，每个残差块每层包含192个卷积核。

学生模型是输入为49个通道，输出为5头部的残差网络，和教师网络相比，增加的2个输出头部是用于软标签数据。神经网络包含4个残差块，每个残差块每层包含128个卷积核。

教师模型网络的参数量为8，693，360个，基准模型的参数量为1，861，552个，学生模型网络的参数量为2，388，186个。由于学生模型的输出包括软标签数据，因此参数量大于基准模型参数量，在进行学生模型和基准模型性能比较时，将学生模型的软标签数据输出头部移除，这样和基准模型一样，都是1，861，552个参数。

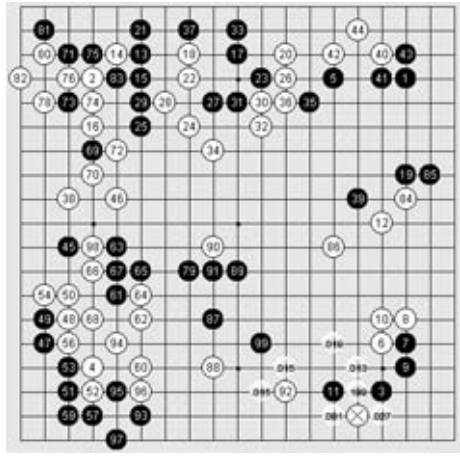

图5 温度为3时的预测值

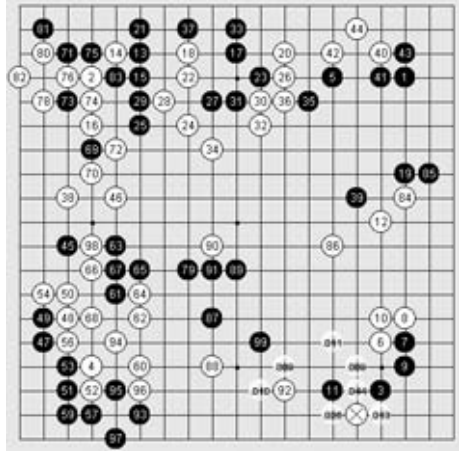

图6 温度为5时的预测值

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

（三）数据集

本课题采用的棋谱是从网络上公开数据集搜集，所有棋谱数据产生于2020年4月之前，然后从中筛选中日韩高水平职业棋手的高质量棋谱。高质量的标准是：著名现代职业棋手、且该棋手属于其国内一流或超一流水准、棋艺顶峰期较长、且能搜集到的该棋手对局数至少在150局以上。

由于AI棋谱和人类棋手棋谱的差距较大。为了探索本课题所采用方法的有效性，以及便于最后生成模型的棋力强弱比较，本课题的棋谱数据剔除了所有AI围棋棋谱，仅包含人类棋手棋谱。最后获取的棋谱总数约为7万余个。

本课题的数据集总量是30，277，632个棋局盘面，训练集大小为27，787，264个棋局盘面，测试集和验证集数据大小各为1，245，184个棋局盘面。

为了加快数据准备时间，部分训练数据利用了围棋棋盘的对称性，即分别通过90度，180度和270度旋转、水平翻转、垂直翻转，共得到8倍数据量。其中约9，830，400个数据是原始棋谱盘面数据，另外17，956，864个数据是由2，244，608个原始棋谱盘面数据通过对称性得到。

为了使测试结果和验证结果更准确，测试集和验证集的所有数据都是原始棋谱盘面数据，不采用通过对称性得到的数据。

在训练得到尧弈教师模型后，在进行知识蒸馏时，我们从训练数据中随机抽取约15%的数据量，共4，194，304个棋局盘面进行训练。测试集和验证集采用和训练教师模型时相同的数据集，仍为各1，245，184个棋局盘面。

（四）训练方法

教师模型训练过程的参数是：训练数据批大小取128，优化函数设定为Adam优化器，学习率设为0.0001，训练10个迭代；然后训练数据集减半，学习率降低为0.00001，再训练3个迭代；最后训练数据集再减半，学习率降低为0.000001，再训练3个迭代。训练总共进行16个迭代后结束。

教师模型在训练集和测试集的下一步走子预测准确率、下下一步走子预测准确率、盘面价值预测准确率如图7至图9所示。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

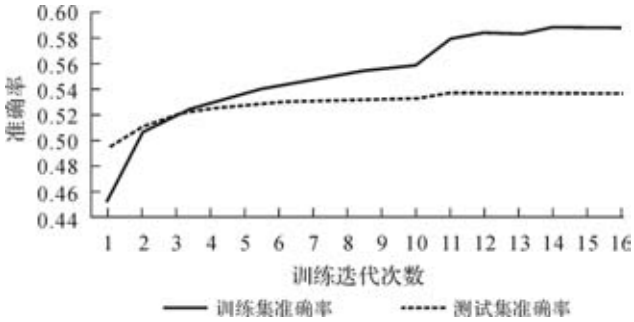

图7　教师模型下一步走子预测准确率

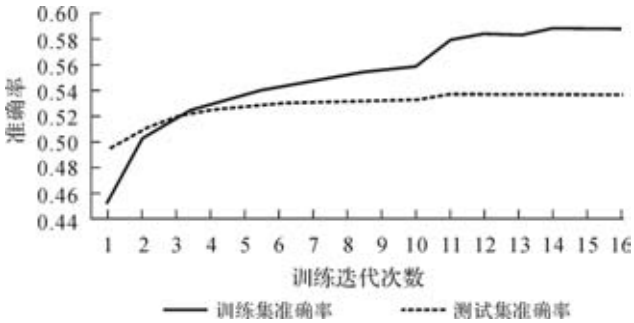

图8　教师模型下下一步走子预测准确率

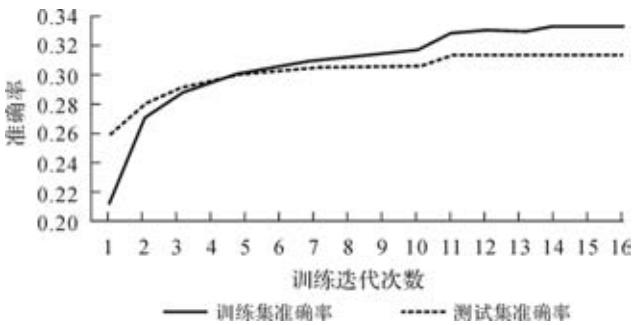

图9　教师模型盘面价值预测准确率

从图7　9中可以看到，当学习率为0.0001时，到第十个训练迭代时，在测试集上预测准确率已不容易提升，但将学习率从0.0001减少为0.00001后，第11个迭代比第10个迭代，在下一步走子、下下一步走子、盘面价值预测准确率等方面，都有明显提升。到第13个训练迭代，在测试集上预测准确率的提升又变得平缓，当将学习率进一步减少为0.000001时，预测准确率的提升已变得不明显。

在训练得到教师模型后，开始训练基准模型和学生模型，温度值分别取2、

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

2.5、3、4、5，共训练5个学生模型。基准模型和学生模型训练过程的参数是：训练数据批大小取128，优化函数设定为Adam优化器，学习率设置为0.0001，基准模型和每个学生模型都训练15个迭代。

图10是基准模型和通过知识蒸馏得到的学生模型在测试数据集上对棋局盘面的下一步走子预测准确率比较。

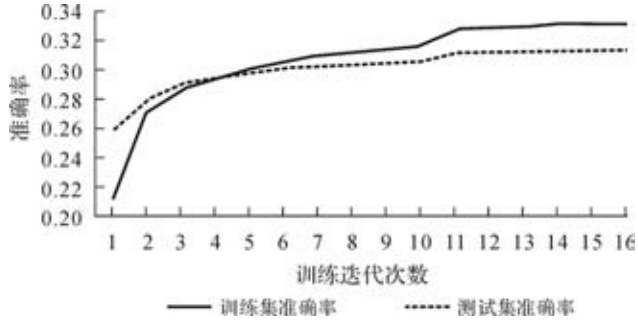

图10　知识蒸馏模型和基准模型下一步走子预测准确率比较

从图10中可以看出，各个知识蒸馏学生模型在测试集上的预测准确率上都比基准模型要好，但不同温度对预测准确率有一定的影响，在温度为2时，预测准确率最好，随着温度的提高，预测准确率逐渐降低。

（五）实验结果评估

对于知识蒸馏方法，本课题采用尧弈非蒸馏模型作为基准模型，比较各个学生模型和基准模型的对弈结果，并按BayesElo算法计算各个模型之间的相对等级分。

在进行对弈时，不进行MCTS模拟，直接根据策略模型走子，每步选择最大的策略值。Google DeepMind在研发早期AI围棋版本时，也是采用这种根据策略模型直接对弈方法和GnuGo 3.8进行棋力比较，后来在AlphaGo中也采用这种方法和当时棋力较强的开源AI围棋软件Pachi进行棋力比较。这种评估方法可以节省时间，且是可行的。

在每个学生模型产生的15个权重模型中，取其中策略值预测准确率最高的模型进行评估。即温度为2、2.5、5的蒸馏模型取第15个迭代模型，温度为3、4的蒸馏模型取第14个迭代模型。

在评估通过知识蒸馏得到的模型和非知识蒸馏模型时，双方对弈的前6步采用随机走子，为了防止开局随机走子下出明显不合理的位置，限定随机走子点在棋

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

盘3　5线或15　17线上。从第7步开始，走子根据各模型的策略网络进行。

各个模型之间各对弈200局，分别执黑100局和执白100局。各个学生模型和基准模型之间的对弈结果如表1所示。

【表1】各个学生模型和基准模型相互对弈结果

	Base	T2	T2.5	T3	T4	T5
Base	-	-	-	-	-	-
T2	144:56	-	-	-	-	-
T2.5	141:59	104:96	-	-	-	-
T3	139:61	90:110	92:108	-	-	-
T4	142:58	87:113	81:119	87:113	-	-
T5	130:70	71:129	66:134	72:128	91:109	-

其中Base表示尧弈基准模型，T2、T2.5、T3、T4、T5分别表示当温度为2、2.5、3、4、5时，通过知识蒸馏得到的学生模型。表中的数据，冒号前数字表示行对手的胜局数，冒号后数字表示列对手的胜局数，例如行T2，列Base的表格数据是144:56，表示T2在总共200局对局中胜了144局，Base胜了56局。

从表1中可以看出，温度为2的知识蒸馏学生模型对基准模型的对弈战绩最好，从图10发现，温度为2的知识蒸馏学生模型，在预测下一步走子的准确率上也是最好的。

根据各个模型之间的对弈结果，我们根据BayesElo算法计算出各个模型的ELO相对等级分，如表2所示。

【表2】各个学生模型和基准模型ELO相对等级分

序号	模型名称	ELO等级分
1	T2	76
2	T2.5	74
3	T3	47
4	T4	5
5	T5	-52
6	Base	-150

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

即知识蒸馏模型和基准模型相比，最高可以提升ELO等级分226分。

从知识蒸馏的角度分析，在不同温度下，学生模型从教师模型中学到的知识量并不相同，有一个最佳的温度点，在这个温度下，学生模型学到的知识最多，棋力也最强。温度太高和温度太低都不好。学到的知识中，除了体现在输出部分对下一步走子点的预测外，也包含了其他知识，这些知识是对输入的总体理解，即特征抽取。所有学到的知识保存在神经网络的中间隐藏层及输出层，而策略值输出只是体现了学到的一部分知识。

六、进一步的工作

由于时间及硬件条件限制，本课题没有对AI围棋的所有优化技术进行实验验证。理论分析可能有效，或已初步得到应用，将来可以进一步探索和进行实验验证的优化技术包括以下几个方面。

（一）采用新神经网络模型架构

本课题所采用的知识蒸馏方法，其中教师模型和学生模型都采用了残差网络架构。当前AI技术发展非常快，新的、更好性能的模型架构层出不穷，可以考虑采用其它网络架构，如轻量级网络MobileNet，EfficientNet，Xception等进行实验验证。

另外，可以在残差网络中增加能扩大全局视野域的构件。因卷积核观察到的只有局部视野域，在网络中增加全局池化层、采用通道相关（channel_wise）的网络等可以提高全局视野域，有助于增加模型棋力。

（二）尝试不同的蒸馏方式

知识蒸馏根据蒸馏算法、模型架构、训练数据类型、知识类型、教师模型和学生模型的相互关系等等不同，有多种蒸馏方式。本课题所采用的知识蒸馏方法是基于响应（response based）的知识蒸馏，将来可以考虑尝试基于特征（feature based）和基于关系（relation based）的知识蒸馏方法。另外本课题的蒸馏过程是采用离线蒸馏方式，将来可以考虑尝试在线蒸馏方式和自蒸馏（self distillation）方式。本课题中的教师模型只有一个，将来可以让学生模型从多个教师模型学习，即同时训练多个模型，然后对多个模型一起蒸馏得到学生模型。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

（三）采用混合压缩技术

在采用知识蒸馏时，并不排斥同时使用其他技术。例如，将知识蒸馏和网络量化结合。网络量化可以看作是一种参数裁剪的压缩技术，在知识蒸馏时采用网络量化，是一个很有应用前景的研究方向。

（四）增加和围棋领域相关的输入特征

增加输入通道，描述更多棋局盘面特征可以提高棋力。例如，征子特征输入采用三个通道比一个通道会更有效，但输入通道过多会相应增加模型的计算负担，并影响MCTS模拟时的速度，因此如何选取输入通道特征需要仔细权衡。

（五）新的学习方式

众所周知，AI围棋在棋力上已远超人类棋手，但AI围棋和人类的下棋方式有明显不同之处，首先AI围棋学习需要大量的棋谱数据，要比人类学棋采用的棋谱多几个数量级；其次，某些概念，如征子对人类来说很容易学习，通过几个简单例子即可掌握，但对AI围棋来说，如果完全通过自对弈强化学习的方法，需要很长学习时间，且最终有可能并没有完全掌握[4]，这可能和卷积神经网络缺乏正确的归纳机制，而只有记忆特定棋型能力有关。也就是说，目前AI围棋所采用的神经网络模型，在模型方面存在缺陷，将来可以探索将新的AI技术应用于围棋领域。

七、结论

本课题对AI围棋的各种优化技术进行了综合分析，重点探讨了知识蒸馏技术，并进行实验验证知识蒸馏是一种实用性强的优化技术。通过知识蒸馏得到的学生模型在模型预测准确率和棋力上有明显的提升，可以有效提升学生模型的性能。在相同配置情况下，棋力最高可以提高ELO等级分226分。

本课题所采用的优化技术除了可应用于围棋领域外，也可以应用其他棋类，如中国象棋、国际象棋等。

另外，AI围棋所采用的优化技术也可以推广应用于其他领域，如教育、医疗、航天、军事、文娱等等领域。

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

参考文献

[1] C. Clark and A. Storkey. Teaching deep convolutional neural networks to play Go，论文预印本平台，http://arxiv.org/abs/1412.3409，December 2014.

[2] C. J. Maddison，A. Huang，I. Sutskever，and D. Silver，Move evaluation in Go using deep convolutional neural networks，论文预印本平台，http://arxiv.org/abs/1412.6564，December 2014.

[3] C. Lee，et al. Human vs. Computer Go: Review and Prospect，IEEE Computational Intelligence Magazine，vol.11，no.3，pp.67–72，August 2016.

[4] Tian，Y.，et al. ELF OpenGo: an analysis and open reimplementation of AlphaZero，In Proceedings of the 36th International Conference on Machine Learning，PMLR 97:6244–6253， 2019.

[5] Coulom，R. Whole–history rating: A Bayesian rating system for players of time–varying strength，In International Conference on Computers and Games，pp.113–124. September 2008.

[6] Silver，D.，et al. Mastering the game of go with deep neural networks and tree search. Nature，vol.529，pp.484–489，January 2016.

[7] Silver，D.，et al. Mastering the game of go without human knowledge. Nature，vol.550，pp.354–359，October 2017.

[8] Silver，D.，et al. A general reinforcement learning algorithm that masters chess，shogi，and go through self–play. Science，vol.362，pp.1140 - 1144，December 2018.

[9] Schrittwieser，J.，et al. Mastering Atari，Go，chess and shogi by planning with a learned model. Nature， Vol.588，pp.604–609，December 2020.

[10] Ciolino，M.，Kalin，J. and Noever，D. The Go Transformer: Natural Language Modeling for Game Play. In 2020 Third International Conference on Artificial Intelligence for Industries，pp.23–26，September 2020.

[11] Wang，H.，et al. Hyper–parameter sweep on AlphaZero general. 论文预印本平台，https://arxiv.org/abs/1903.08129，2019.

[12] Wu，D. J. Accelerating self–play learning in Go. 论文预印本平台，https://arxiv.org/abs/1902.10565，2019.

[13] Wu，T. R.，Wei，T. H.，& Wu，I. C. Accelerating and Improving

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

AlphaZero Using Population Based Training. In Proceedings of the AAAI Conference on Artificial Intelligence，Vol.34，No.01，pp.1046−1053，April 2020.

[14] Chen，Y. C.，Chen，C. H.，& Lin，S. S. Exact−win strategy for overcoming AlphaZero. In Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems， pp.26−31，November 2018.

[15] Kadam，P.，Xu，R. and Lieberherr，K. Dual Monte Carlo Tree Search. 论文预印本平台，https://arxiv.org/abs/2103.11517，2021.

[16] C. Bucilua，R. Caruana，and A. Niculescu−Mizil. Model compression，in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining，pp.535－541，August 2006.

[17] G. Hinton，O. Vinyals，and J. Dean，Distilling the knowledge in a neural network，论文预印本平台，https://arxiv.org/abs/1503.02531.

[18] J. P. Gou，et al. Knowledge Distillation: A Survey. 论文预印本平台，https://arxiv.org/abs/2006.05525，2021.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net