# AlphaGo Zero与能力增强 AlphaGo Zero与能力增强 原文链接： 是AI能力的一个很好的演示，也是一个很好的观点证明了一个有效的对应策略。 AlphaGo Zero是如何工作的 AlphaGo Zero 通过两个函数进行学习（作为当前方式的输入） 一个超前的过度动作p是被训练来预测AlphaGo 最终落子位置的。 一个价值函数v是被训练来预测哪一位选手会赢。（如果是AlpahaGo 自己和自己下。） 这两个训练都是监督学习。一旦我们有了这两个函数，AlphaGo 将用p和v去判断Monte Carlo树的1600步得到的结果。通过这样高代价的搜索过程来训练p，使p最后有一个好的移动表现。随着p的增强，高代价的搜索就变得越来越强大，而p就越来越趋向于理想化的那个值。

原文链接：[AlphaGo Zero and capability amplification](https://ai-alignment.com/alphago-zero-and-capability-amplification-ede767bb8446?from=hackcv&hmsr=hackcv.com&utm_medium=hackcv.com&utm_source=hackcv.com)

[AlphaGo Zero](https://deepmind.com/blog/alphago-zero-learning-scratch/) 是AI能力的一个很好的演示，也是一个很好的观点证明了一个有效的对应策略。

### AlphaGo Zero是如何工作的

AlphaGo Zero 通过两个函数进行学习（作为当前方式的输入）

* 一个超前的过度动作p是被训练来预测AlphaGo 最终落子位置的。
* 一个价值函数v是被训练来预测哪一位选手会赢。（如果是AlpahaGo 自己和自己下。）

这两个训练都是监督学习。一旦我们有了这两个函数，AlphaGo 将用p和v去判断Monte Carlo树的1600步得到的结果。通过这样高代价的搜索过程来训练p，使p最后有一个好的移动表现。随着p的增强，高代价的搜索就变得越来越强大，而p就越来越趋向于理想化的那个值。

### 迭代能力增强 迭代能力增强 在中,我们训练了一个函数： 一个“弱”策略A，它被训练来预测在给定的情况下代理最终会决定什么。 就像AlphaGo 不会用p直接决定移动，我们也不会用弱策略直接决定行动。我们会用能力增强的方案替代：我们多次调用A以取得更加智能的判断。我们训练A去忽视高代价的增强过程，为了直接取得智能的判断。随着A的增强，放大的策略变得更加强大，A不断趋向于理想化的那个值。 以AlphaGo Zero举例，A是超前的过度动作，增强策略是MCTS。（确切的说，A是一对（p,v)，而能力增强厕所是MCTS加上使用一个滚动查看谁会赢。）

在[迭代能力增强最简单的形式](https://ai-alignment.com/benign-model-free-rl-4aae8c97e385)中,我们训练了一个函数：

* 一个“弱”策略A，它被训练来预测在给定的情况下代理最终会决定什么。

就像AlphaGo 不会用p直接决定移动，我们也不会用弱策略直接决定行动。我们会用能力增强的方案替代：我们多次调用A以取得更加智能的判断。我们训练A去忽视高代价的增强过程，为了直接取得智能的判断。随着A的增强，放大的策略变得更加强大，A不断趋向于理想化的那个值。

以AlphaGo Zero举例，A是超前的过度动作，增强策略是MCTS。（确切的说，A是一对（p,v)，而能力增强厕所是MCTS加上使用一个滚动查看谁会赢。）

抛开AlphaGo Zero 不说，A可以看作是一个问题回答系统，它在很多时候被用于将一个问题分解成各个部分后分散解决。又或者可以把它看成是一个认知工作空间的，很多时候可以用于使解决问题时想的更远。  
抛开AlphaGo Zero 不说，A可以看作是一个问题回答系统，它在很多时候被用于将一个问题分解成各个部分后分散解决。又或者可以把它看成是一个认知工作空间的，很多时候可以用于使解决问题时想的更远。  
意义  
加强学习者用激励函数去优化它，但不幸的是，我们并不知道在哪里用激励函数可以使它忠实的追踪我们所关心的事。这是安全问题的一个重要来源。  
相比之下，AlphaGo Zero 采取策略改进操作（比如MCTS）并收敛到它的不动点。如果我们能够找到一个方法去改进这个策略的同时使其不变，那么我们可以应用同样的算法来获得非常强大但整齐的策略。

### 意义

加强学习者用激励函数去优化它，但不幸的是，我们并不知道在哪里用激励函数可以使它忠实的追踪我们所关心的事。这是安全问题的一个重要来源。

相比之下，AlphaGo Zero 采取策略改进操作（比如MCTS）并收敛到它的不动点。如果我们能够找到一个方法去改进这个策略的同时使其不变，那么我们可以应用同样的算法来获得非常强大但整齐的策略。

使用MCTS在现实世界中去实现一个简单的目标都无法使其对齐，因此它不符合要求。但是在“”的方面是可能做到的。只要我们以一个对其的政策开始—“想要“使其对齐，从某种意义上说，允许它思考的远一些就可能使它更聪明也更一致。  
使用MCTS在现实世界中去实现一个简单的目标都无法使其对齐，因此它不符合要求。但是在“”的方面是可能做到的。只要我们以一个对其的政策开始—“想要“使其对齐，从某种意义上说，允许它思考的远一些就可能使它更聪明也更一致。  
我认为在如今设计保持对其的方法增强使一个容易处理的问题，可以在现有ML或人工调整的上下文中对其进行研究。所以我认为这使一个在AI校准方面一个可以一试的方向。后备方案可以直接合并到AlphaGo Zero的体系结构中，所以我们已经可以获得相关经验的反馈。如果运气好的话，表现好的AI系统会表现得像AlphaGo Zero,这就可以给我们很多方法去校准AI。

我认为在如今设计保持对其的方法增强使一个容易处理的问题，可以在现有ML或人工调整的上下文中对其进行研究。所以我认为这使一个在AI校准方面一个可以一试的方向。后备方案可以直接合并到AlphaGo Zero的体系结构中，所以我们已经可以获得相关经验的反馈。如果运气好的话，表现好的AI系统会表现得像AlphaGo Zero,这就可以给我们很多方法去校准AI。