Mastering the Game of Go without Human Knowledge

文献紹介　12 Jul. 2019　鈴木遊

参考論文：

[1] D. Silver, J. Schrittwieser, K. Simonyan, et al., Nature 550, 354-359 (18 Oct. 2017)
[2] D. Silver, J. Schrittwieser, K. Simonyan, et al., Nature 529, 484-489 (28 Jan. 2016).

1997年 5月11日

Garry Kasparov – Deep Blue

1勝（3引き分け）2勝

囲碁・将棋界の反応

余裕？

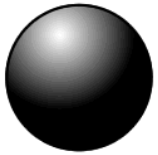
囲碁・将棋という壁

チェス

- 終盤ほど駒数が減っていく
- 駒の損得が重要

将棋

- 取った駒を使うことができる
- 駒の損得がチェスほど重要でない



囲碁

- 盤面が広い
- 形勢や損得が抽象的である

探索局面が多い　局面の評価が難しい

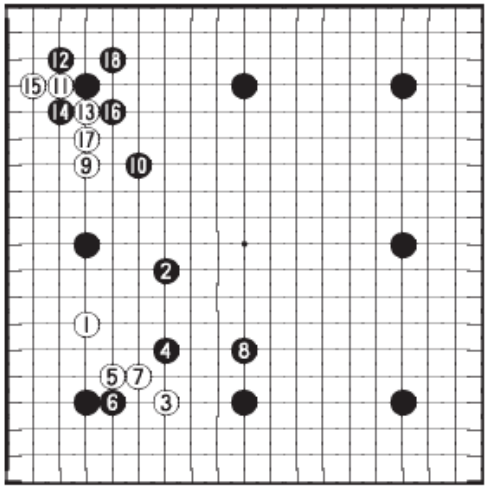
2000年代

囲碁・将棋という壁

[将棋] ○ 渡辺明 竜王 ー ● Bonanza

アマ五段くらい？

[囲碁] ● 青葉かおり四段 ー ○ Crazy Stone (8子局)



アマ二～三段くらい？

2016年 3月15日

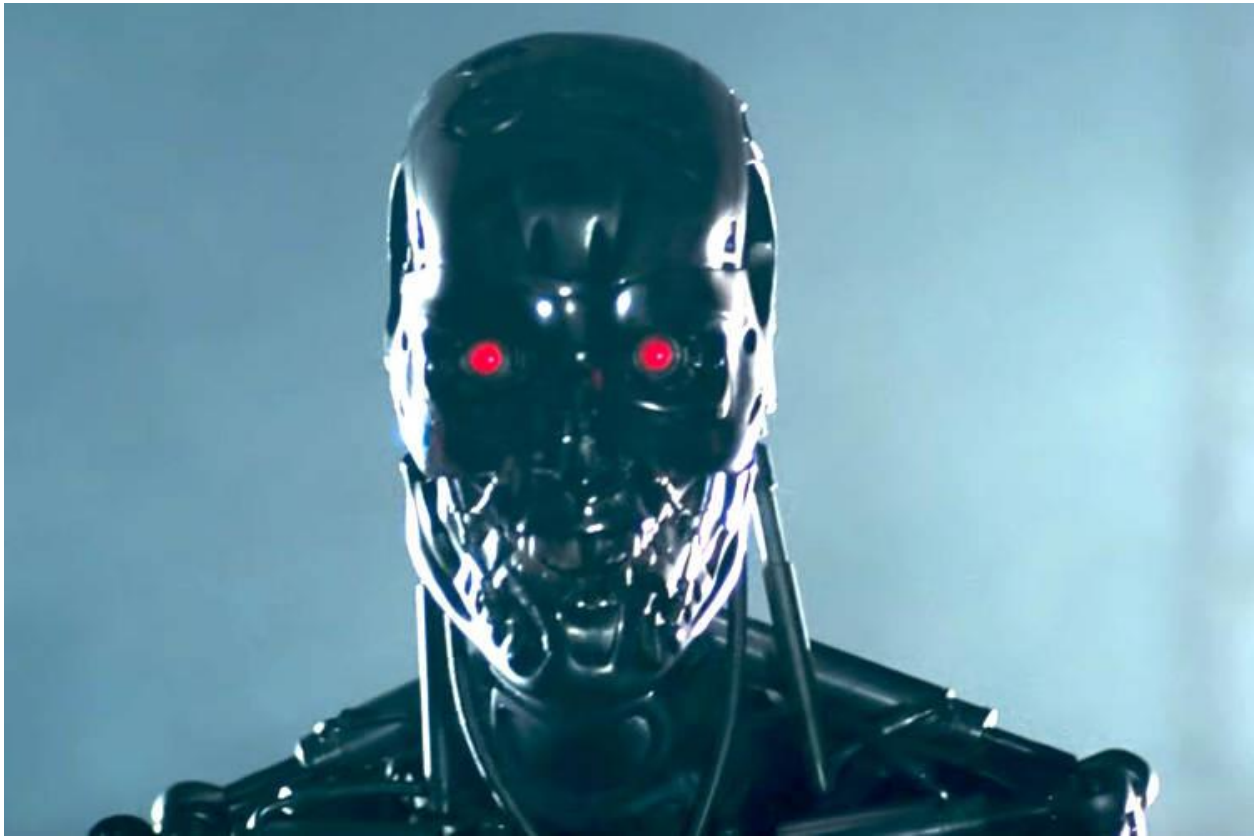
李世ドル九段 – AlphaGo

1勝　　4勝

さらに翌2017年

AlphaGo – AlphaGo zero

０勝　　１００勝

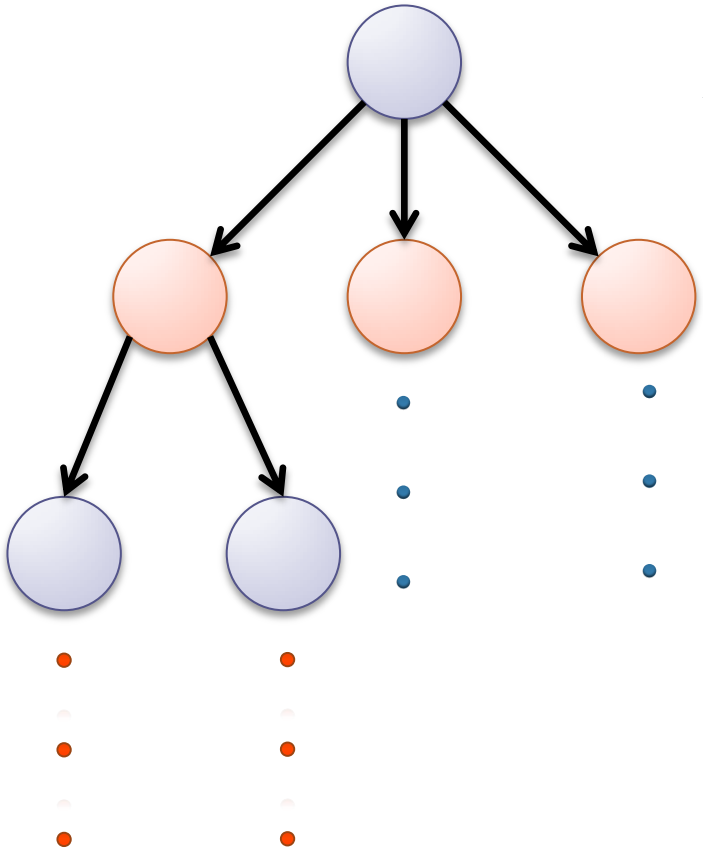


AlphaGo – AlphaGo zero

0勝　　１００勝

20年で何があったのか？

ゲーム木



完全ゲーム木 ～ 完全解析
　通常は不可能
　(チェスなら$O(10^{120})$)

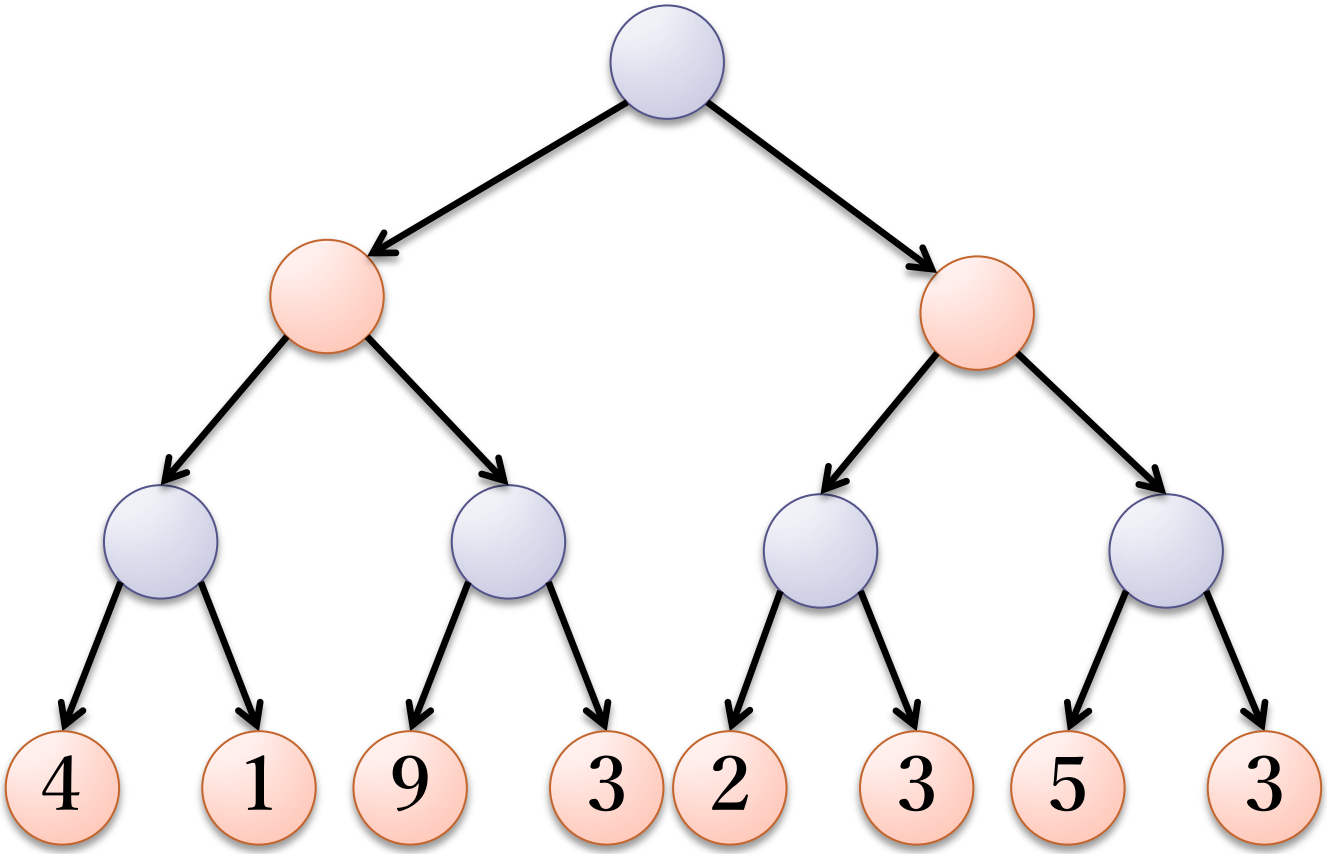
部分ゲーム木
　枝を刈り取ったもの
　質の良い部分ゲーム木が必要

- 無駄な手を読まない
- 勝ち負けの正確な判断

例. 将棋ソフト Bonanza

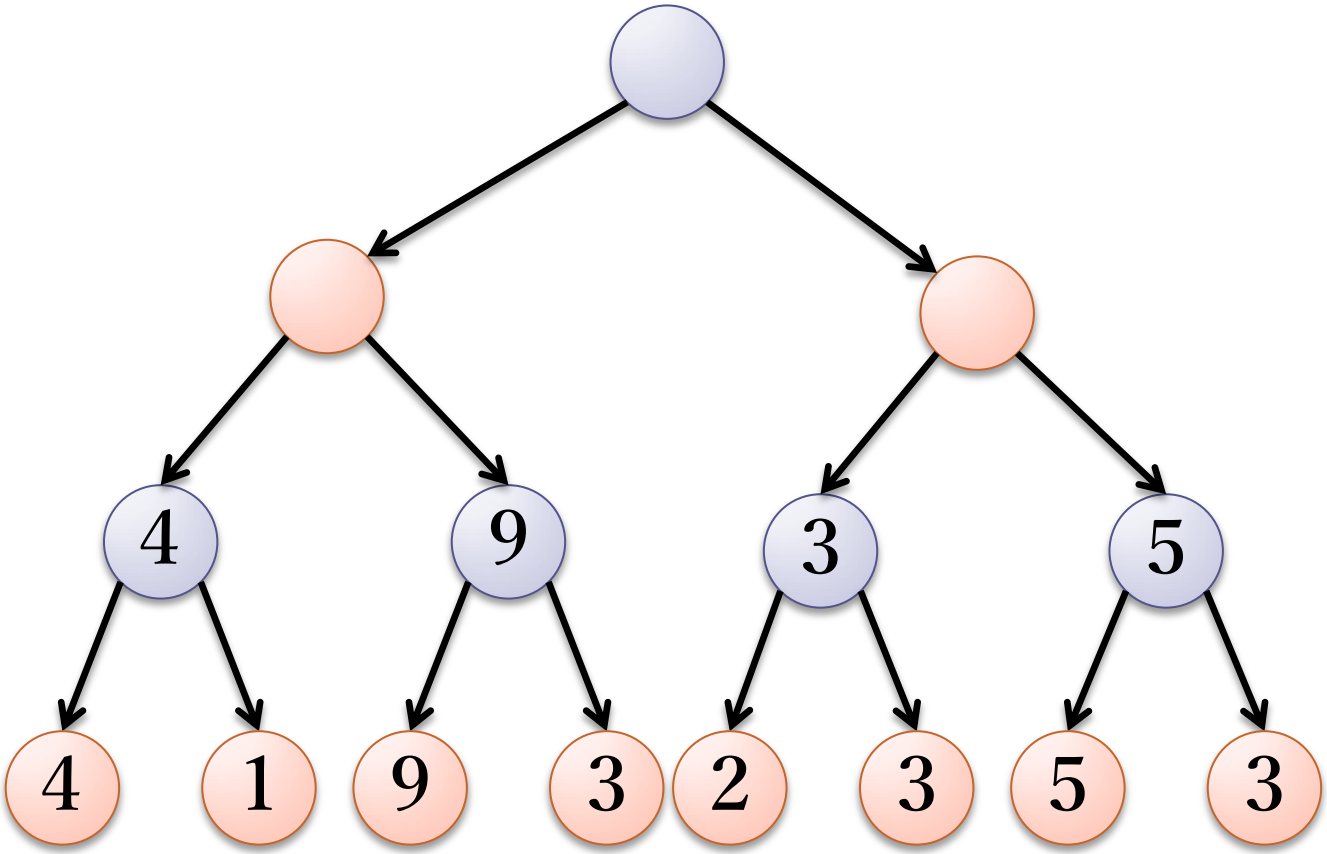
1. 無駄な手を読まない＝探索部

Mini – Max 法



自分は最大,相手は最小
下から上へ

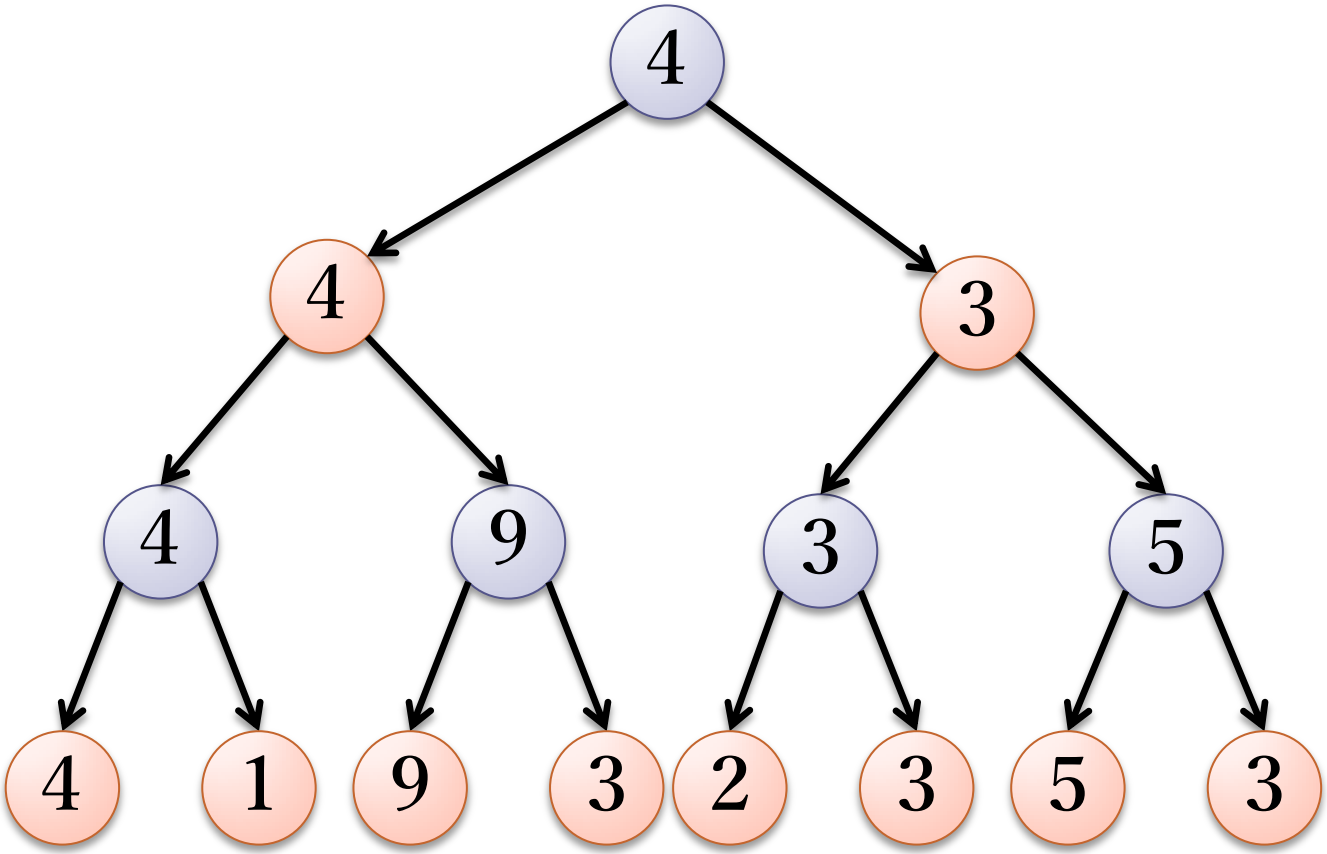
Mini – Max 法



自分は最大,相手は最小

下から上へ

Mini – Max 法



自分

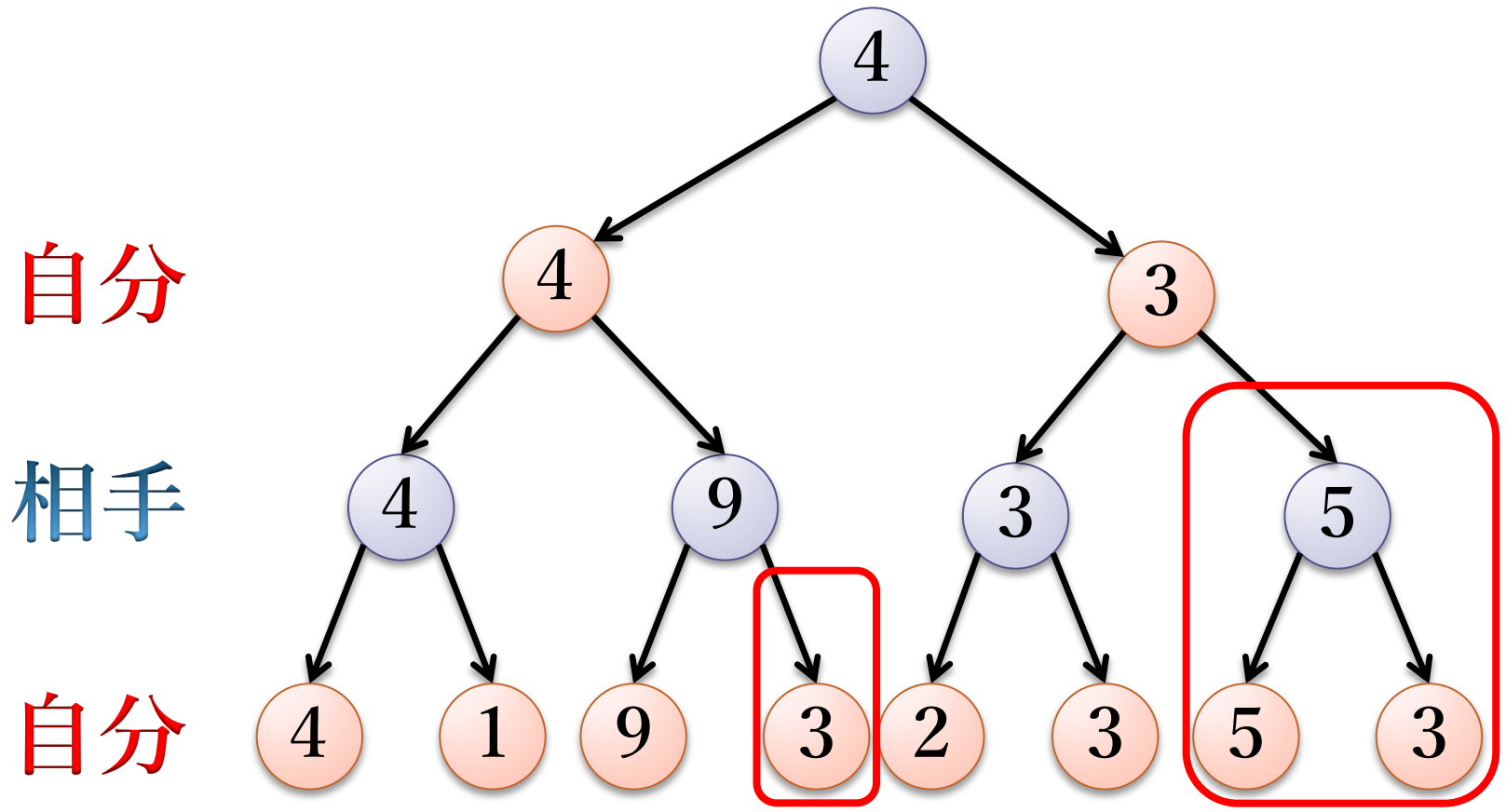
相手

自分

自分は最大,相手は最小

下から上へ

Alpha – Beta 法



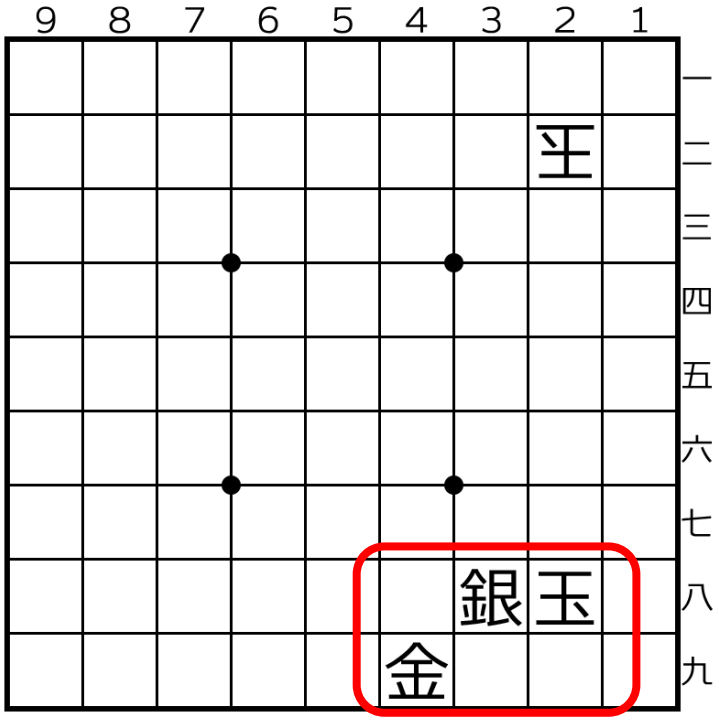
読む必要がない手もある

Bonanza は Null move, Futility を併用

2. 勝ち負けの正確な判断＝評価部

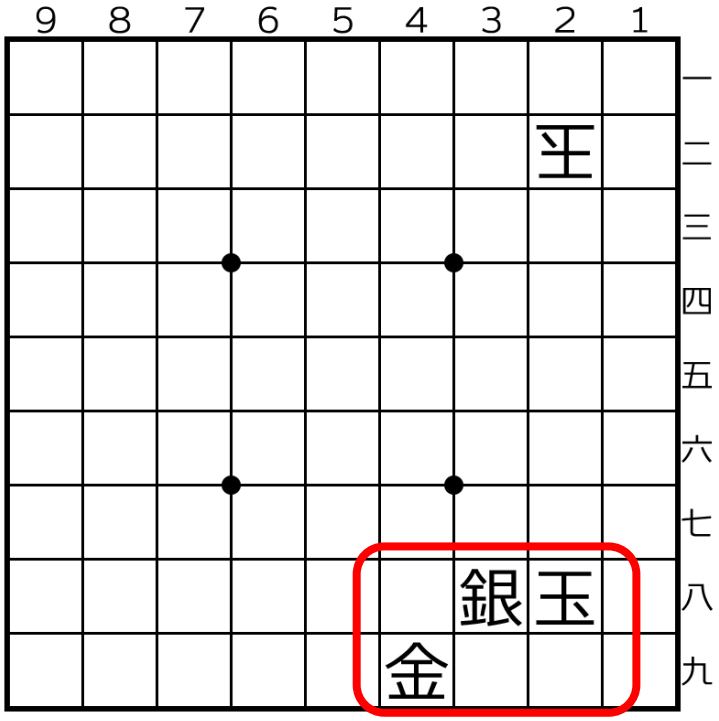
Bonanza型評価関数

- 駒の価値
  - 飛車 ○点, 角行 ○点, ・・・
- 位置関係
  - 王将と任意の駒２つ
- 一万くらいのパラメータを機械学習で調整



Bonanza型評価関数

- 駒の価値
  - 飛車 ○点, 角行 ○点, ・・・
- 位置関係
  - 王将と任意の駒２つ
- 一万くらいのパラメータを機械学習で調整



まとめ

- 全幅(＝力任せ)探索 + 上手い評価関数
- 将棋に有効な fit ansatz が見つかった

囲碁での形成判断

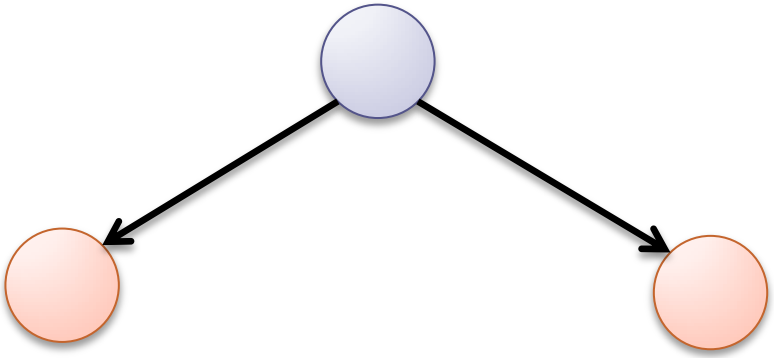

34
28
5
29
2
1
15
18
13
12
16
14
27
33
39
30
31
37
35
32
36
38
24
44
22
46
40
43
20
19
17
42
41
23
21
10
50
47
45
48
8
25
49
6
7
4
9
26
11
3

囲碁での形成判断

囲碁の評価部をどう作るか？

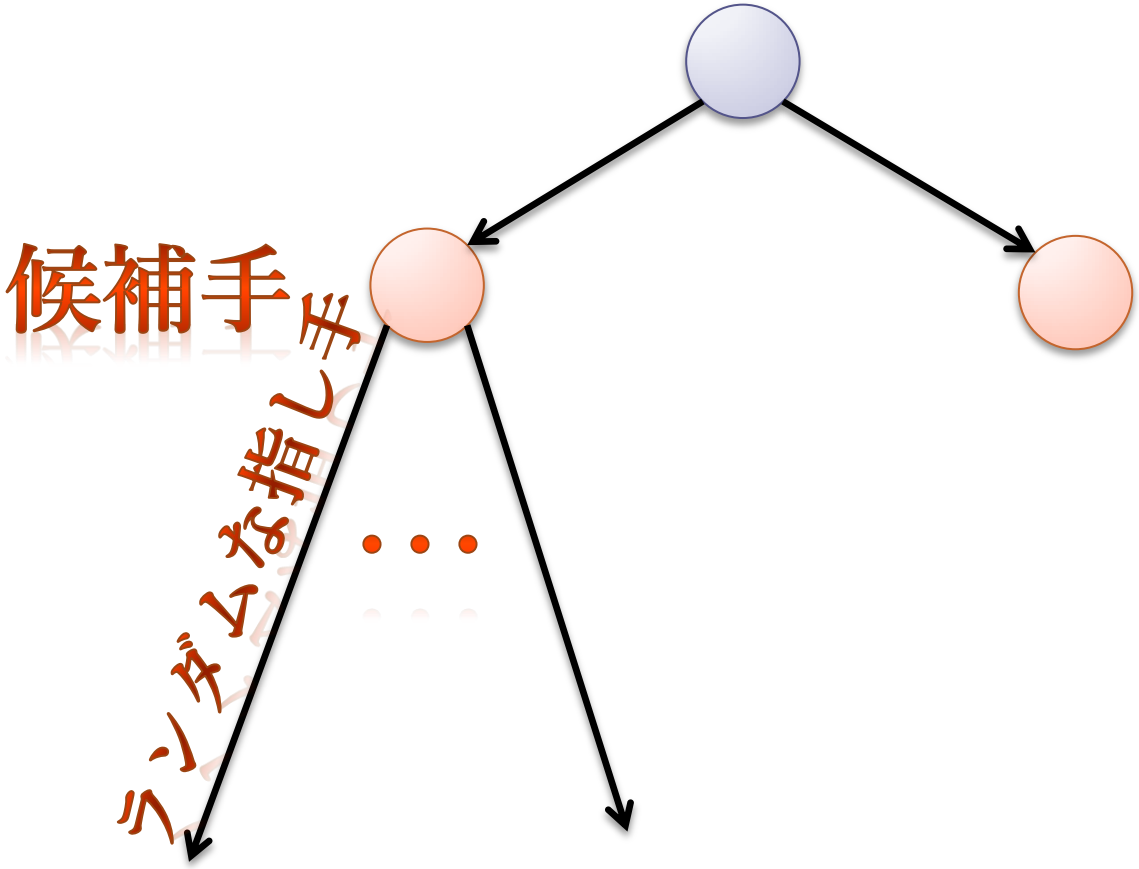
モンテカルロ木探索

候補手



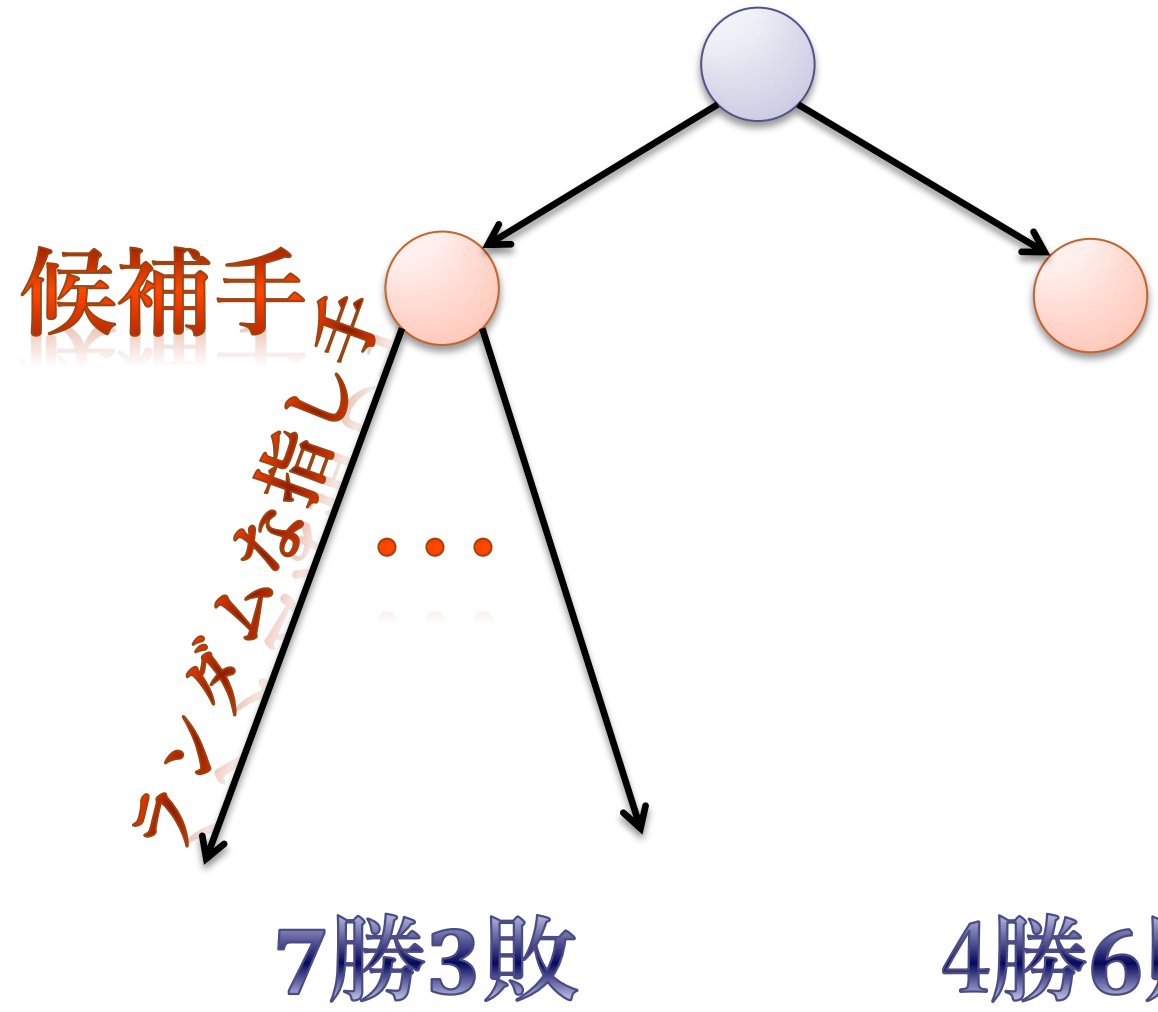
- 候補手の優劣を決めたい

モンテカルロ木探索



- 候補手の優劣を決めたい
- ランダムに終局までプレイ
  - 囲碁に特有

モンテカルロ木探索



- 候補手の優劣を決めたい
- ランダムに終局までプレイ
  - 囲碁に特有
- 勝率が高い手を選ぶ
- 当然弱い
  - MC 対 従来法
  - 勝率 約1割

Multi-Armed Bandit 問題

100/300　100/300　100/300

- スロットで儲けたい(利益を最大化したい)
- 資源を等分しても当然ダメ

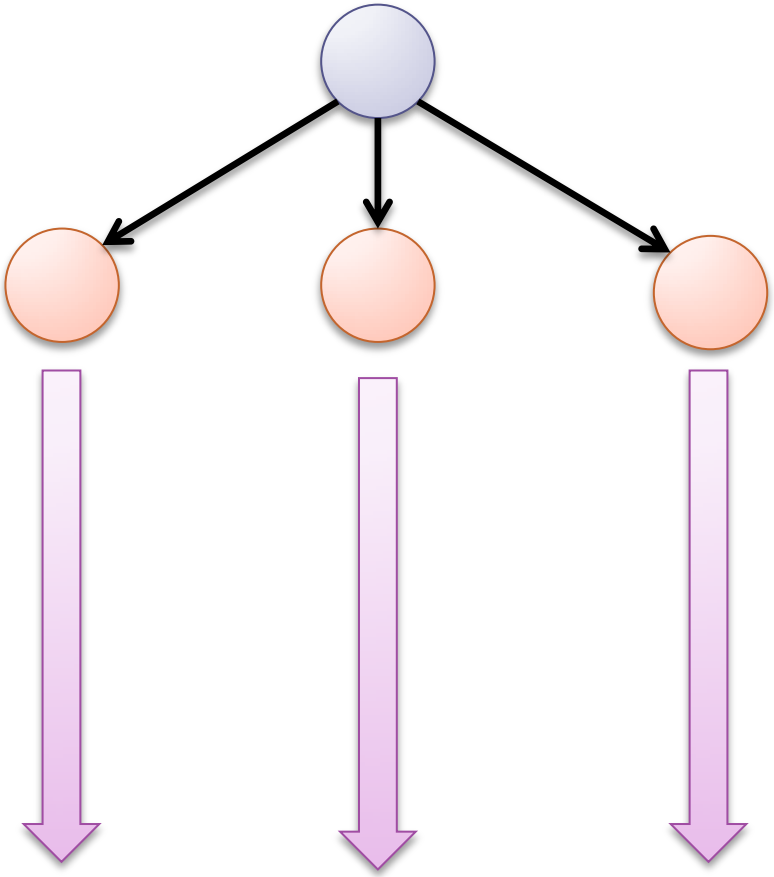
Multi-Armed Bandit 問題

200/300　80/300　20/300

- スロットで儲けたい(利益を最大化したい)
- 資源を等分しても当然ダメ
- 少しプレイして、有望な台に投資

Crazy Stone

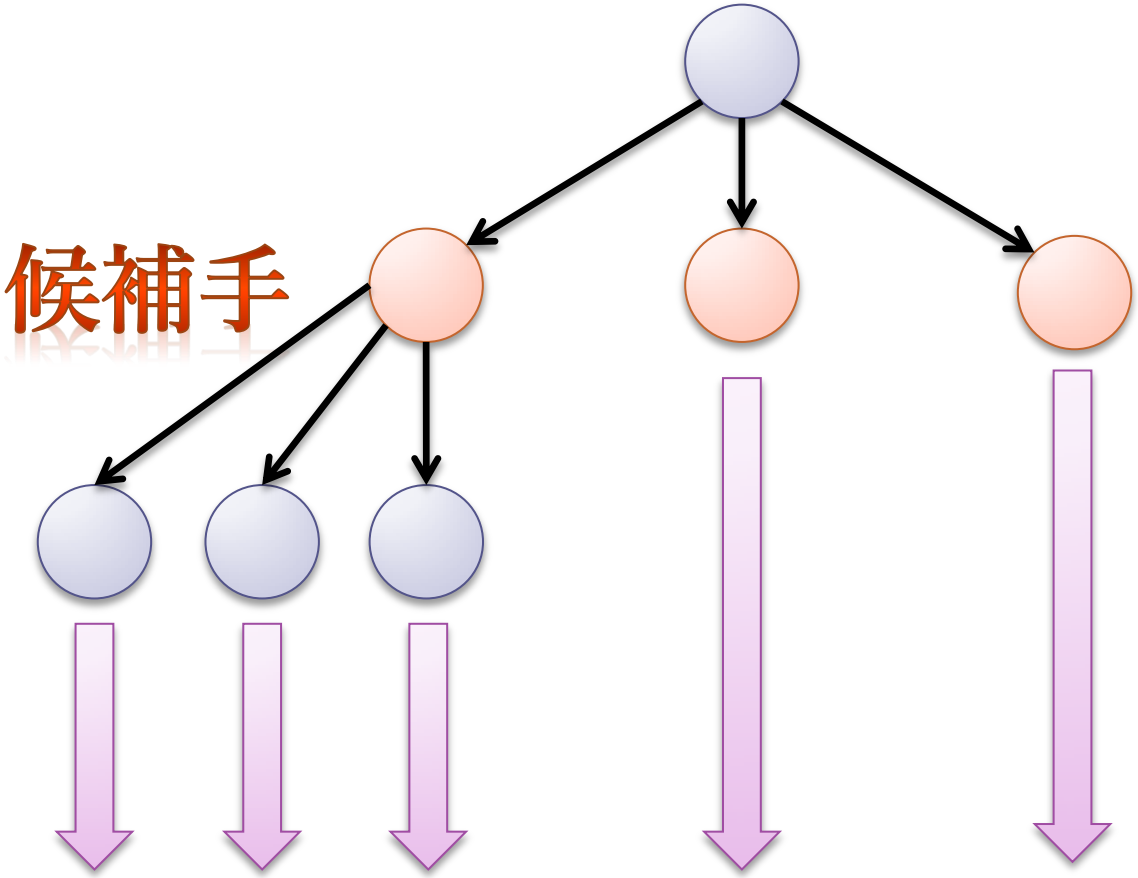
候補手



9勝1敗 3勝7敗 4勝6敗

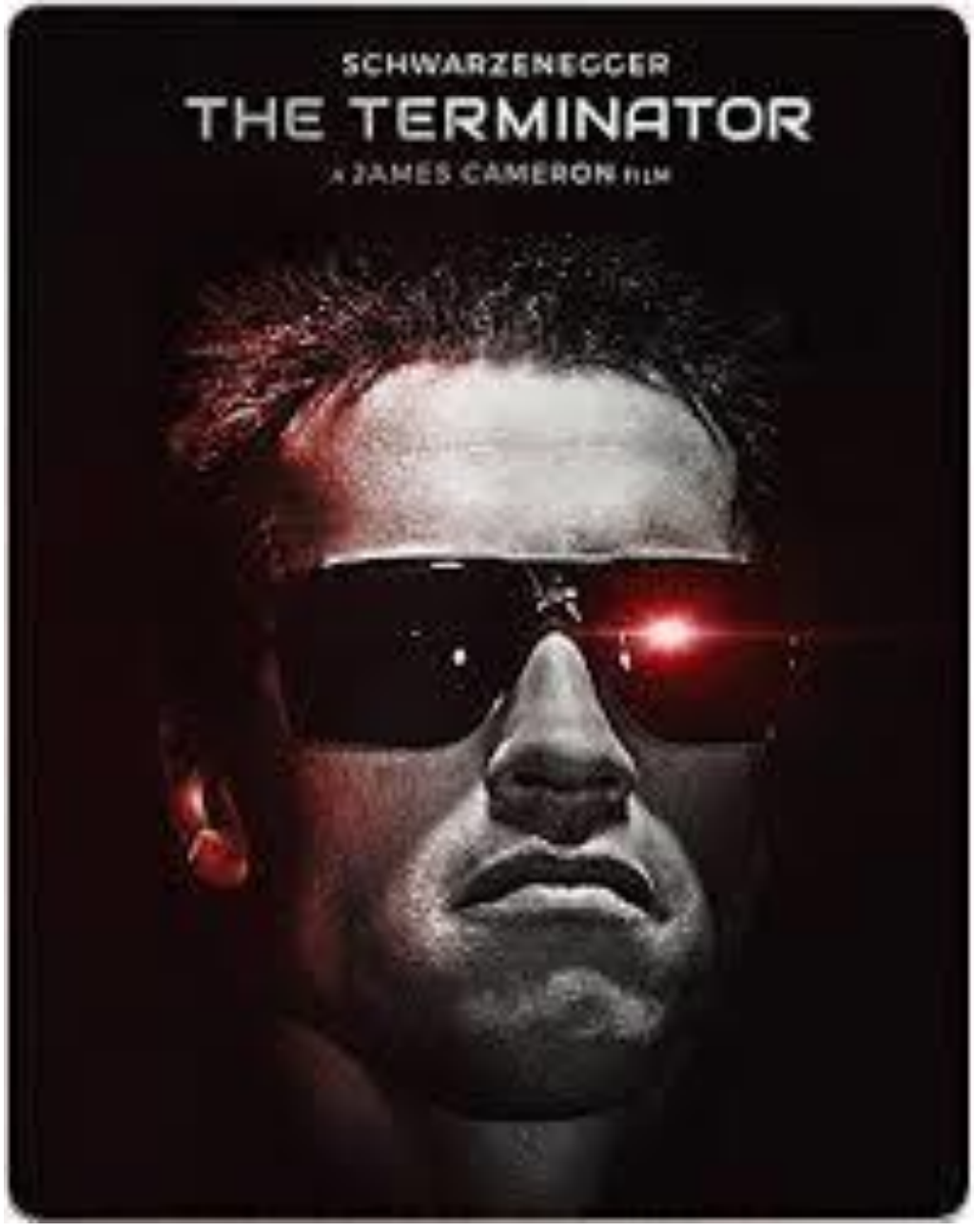
- 最初は平等に投資
- 有望な手
  - 試行回数を増やす
  - 探索を深くする

Crazy Stone

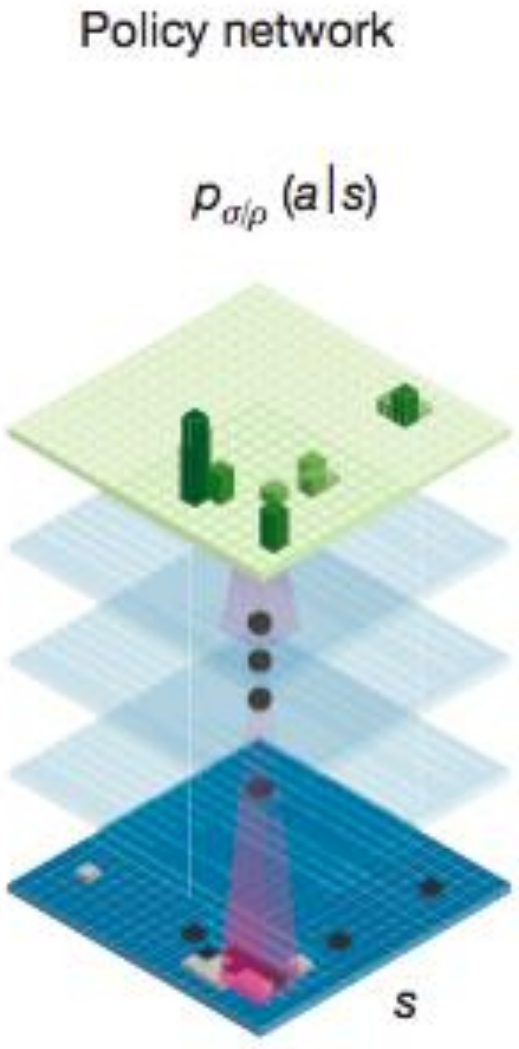


- 最初は平等に投資
- 有望な手
  - 試行回数を増やす
  - 探索を深くする
- 並列処理に適している
- 50,000 node/sec
- 囲碁AIの break-through

AlphaGo の登場

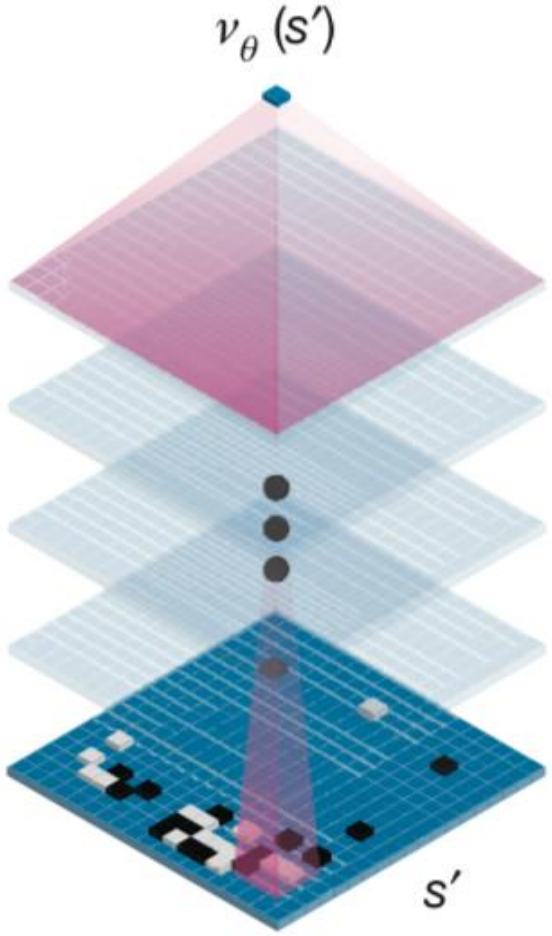

SCHWARZENEGGER
THE TERMINATOR
A JAMES CAMERON FILM

探索部(Policy Network)



- Deep Leaning を用いて
  手の選択確率を計算する
- KGS の棋譜 16万局から学習
- そこから更に強化学習

評価部(Value Network + Rollout Policy)



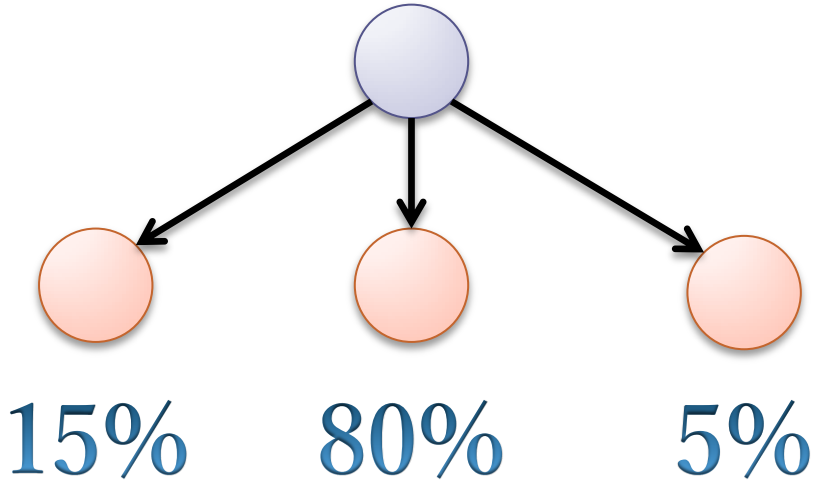
Value Network

- Deep Leaning を用いて
  局面の勝率を計算する
- 学習は KGS + 強化学習

Rollout Policy

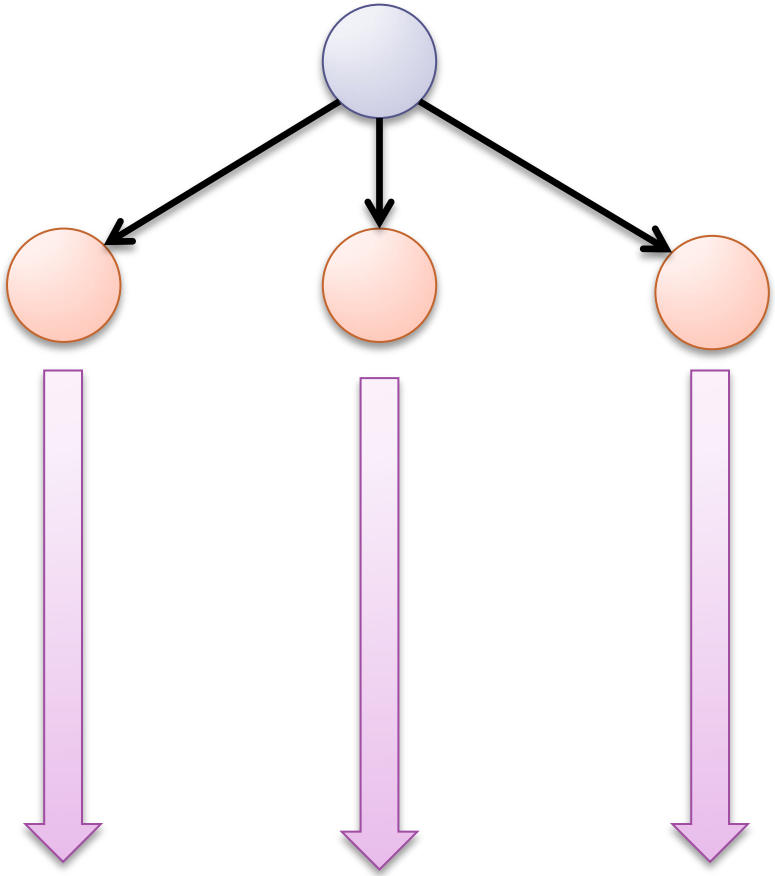
- 簡単、高速化された
  Policy Network
- モンテカルロ木探索に使われる

AlphaGo のモンテカルロ



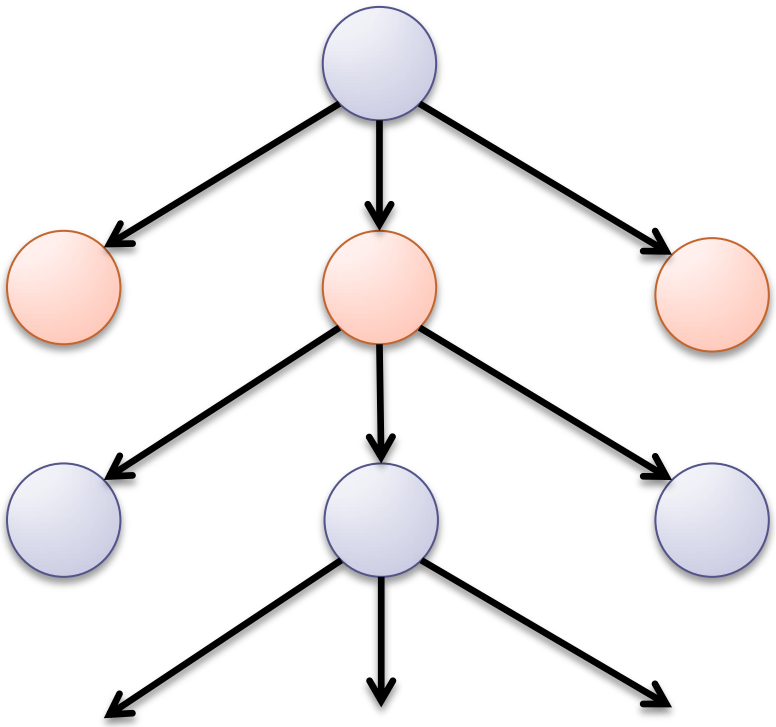
- Policy Network で候補手の選択確率を計算
  - 確率が高い手を詳しく調べる

AlphaGo のモンテカルロ



- Policy Network で候補手の選択確率を計算
  - 確率が高い手を詳しく調べる
- Value Network で勝率を計算
- Rollout Policy でもっともらしい終局へ

AlphaGo Zero



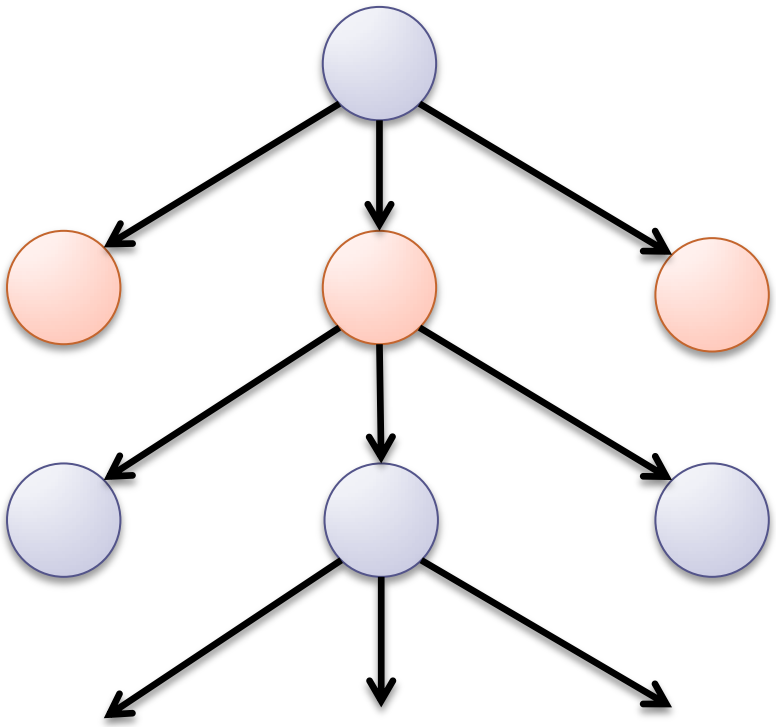
- Policy Net. と Value Net. を統合(Dual Net.)
- Rollout は廃止
- 人間の棋譜は使わない
  - 初期はランダム
- 入力の減少

石の配置
手の履歴
呼吸点の数
シチョウ　など

→

石の配置
手の履歴
手番

AlphaGo Zero



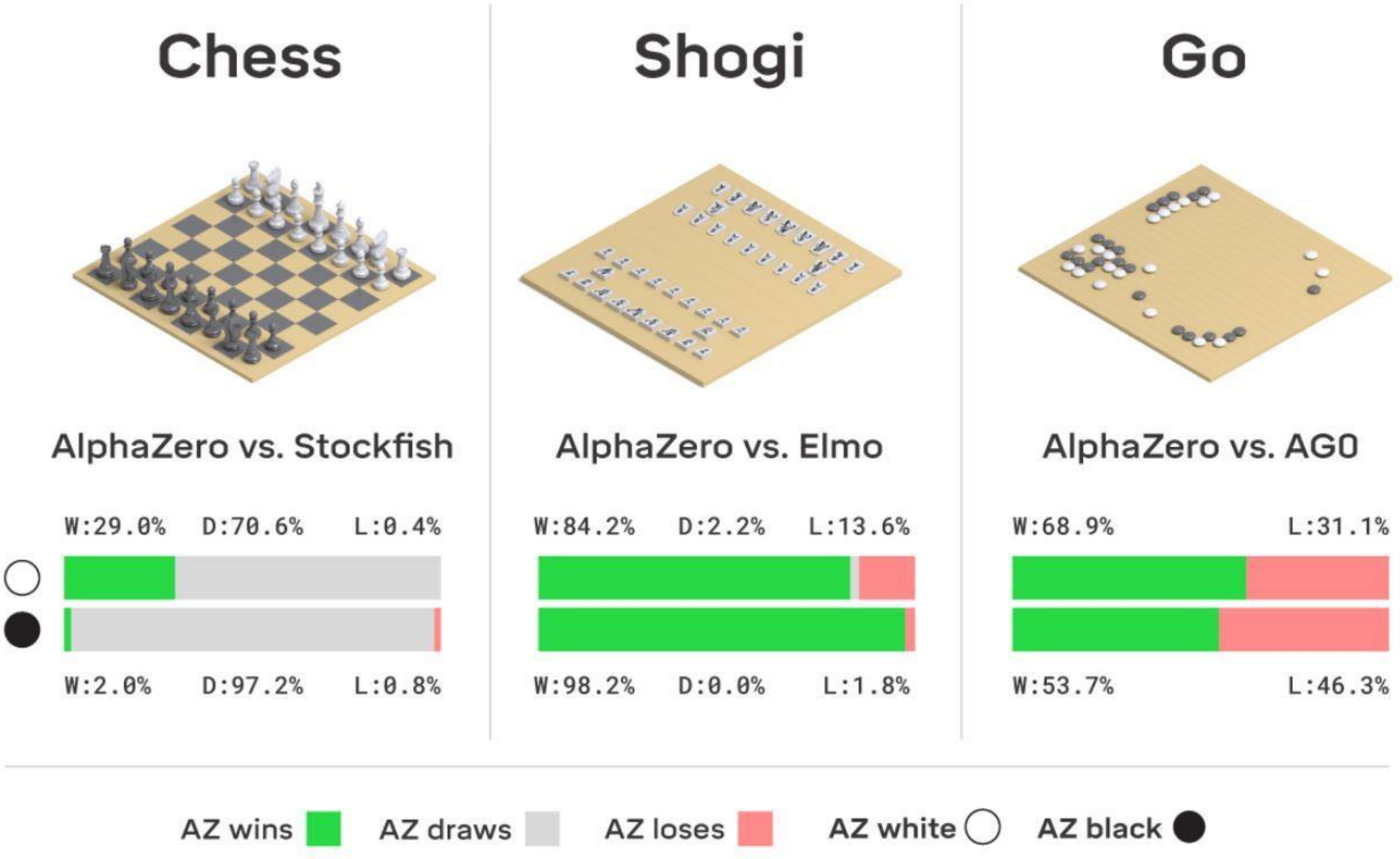
- Policy Net. と Value Net. を統合(Dual Net.)
- Rollout は廃止
- 人間の棋譜は使わない
  - 初期はランダム
- 入力の減少

石の配置
手の履歴
呼吸点の数
シチョウ　など

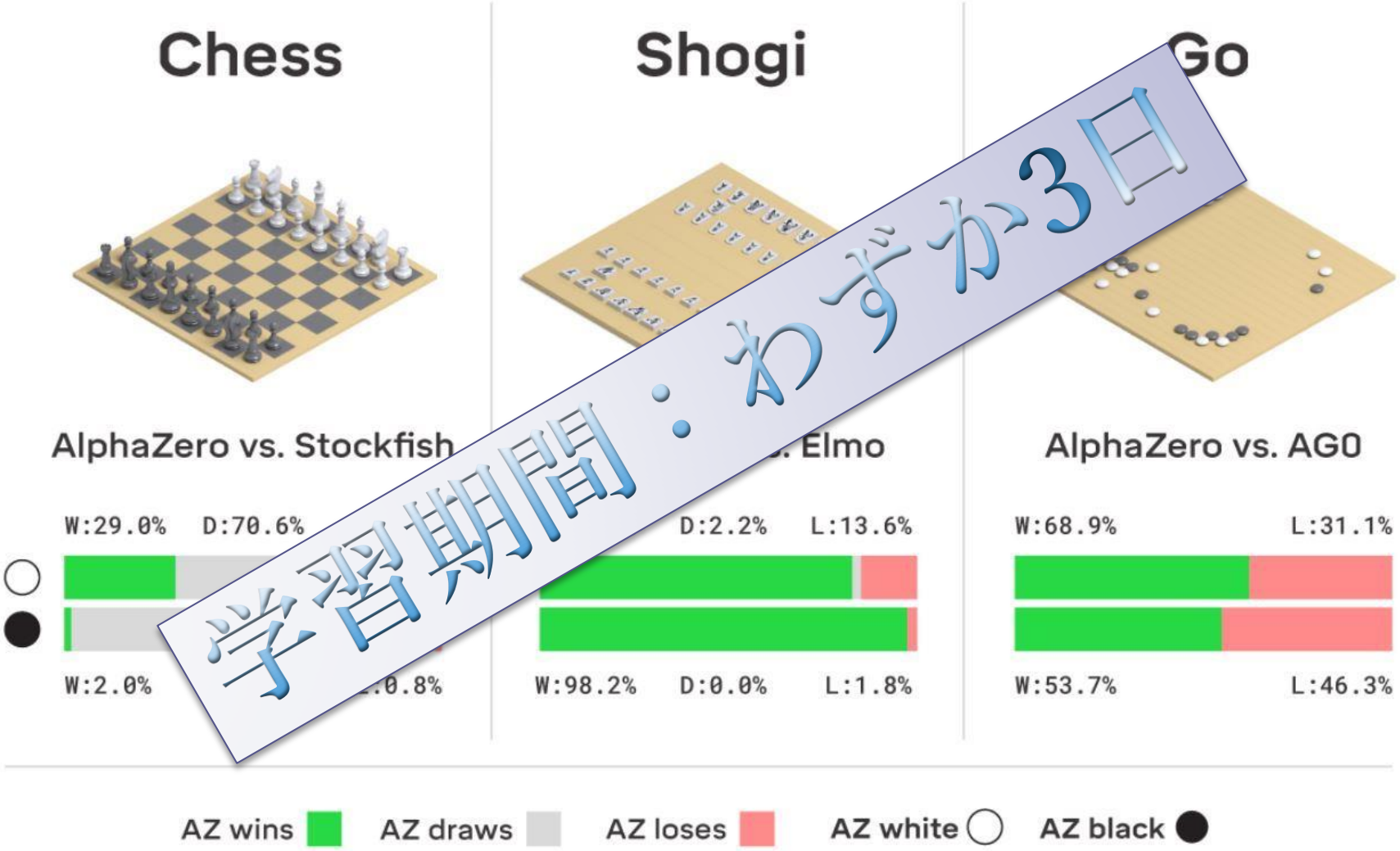
→ 石の配置
手の履歴
手番

1.人間の知識を使わない 2.囲碁に特化していない

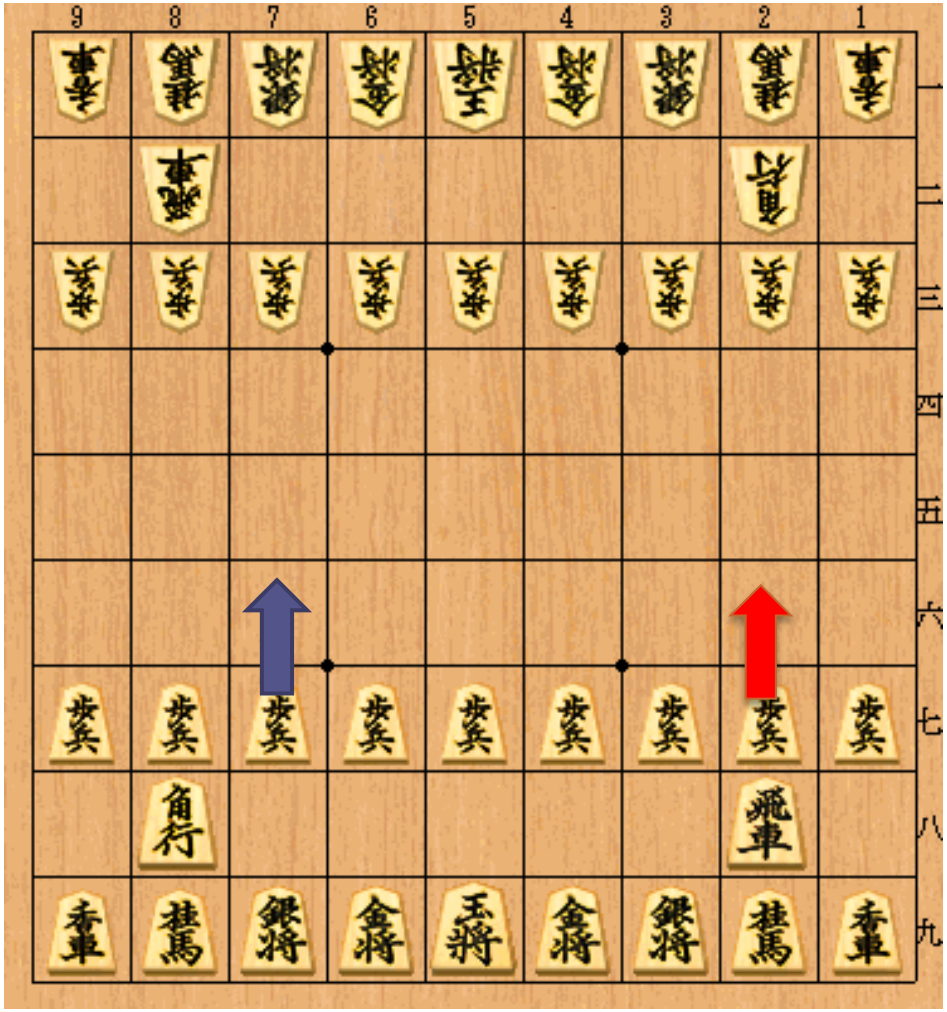
チェス, 将棋, 囲碁を同じアルゴリズムで



チェス，将棋，囲碁を同じアルゴリズムで



個人的に面白かった点



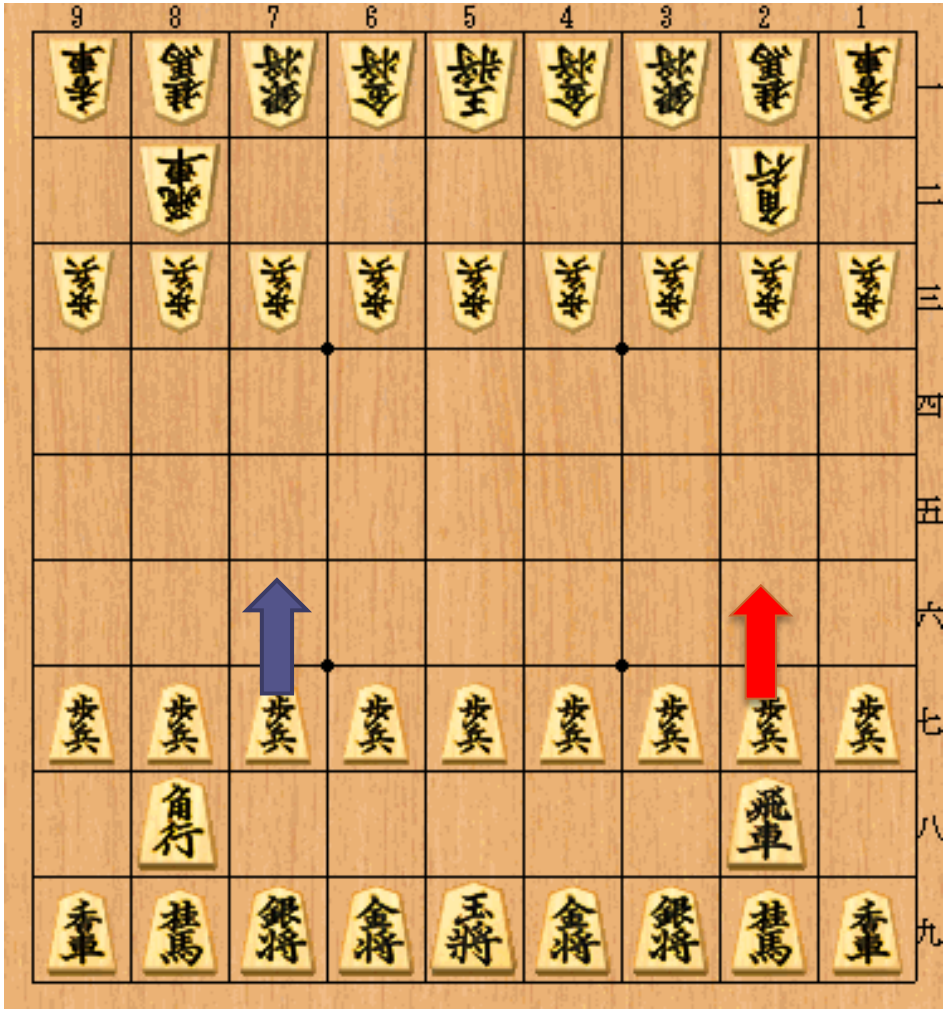
７六歩

- 振り飛車
- 矢倉
- 横歩取り
- 角換わり

２六歩

- 横歩取り
- 角換わり
- 相掛かり

個人的に面白かった点



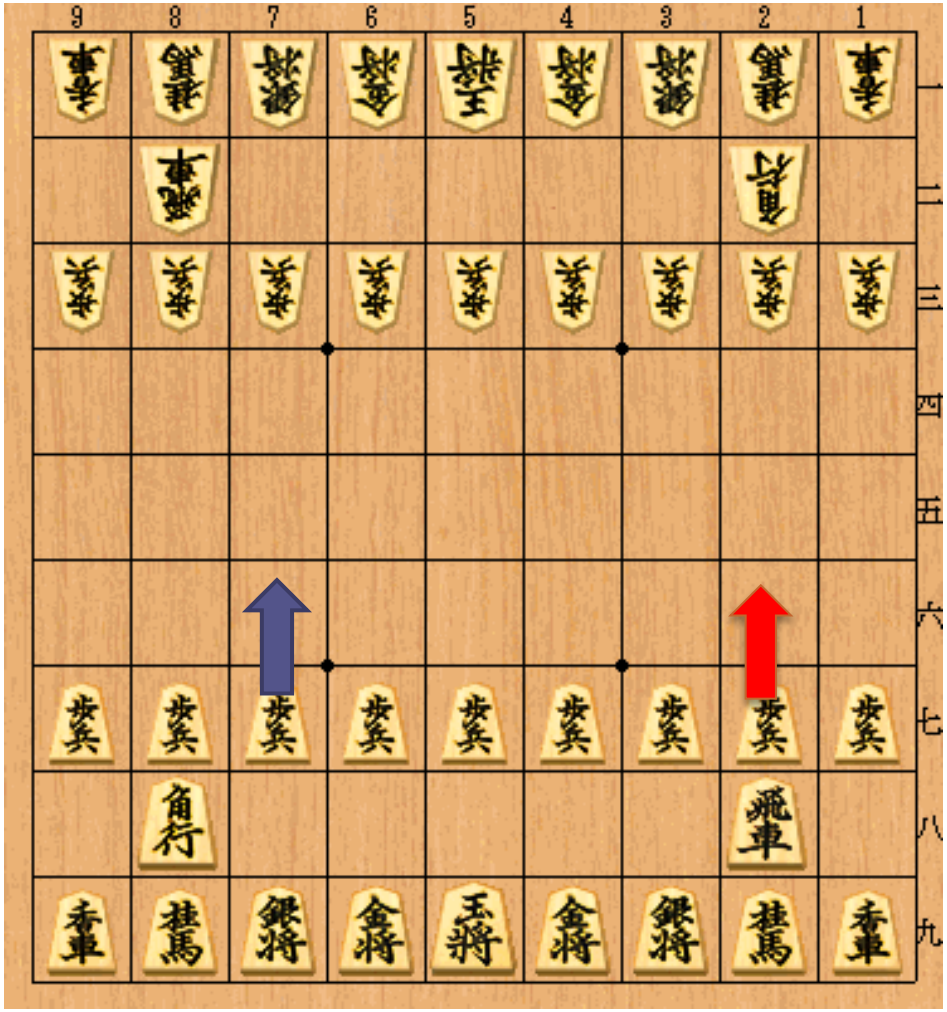
~~7六歩~~　　2六歩

~~振り飛車~~
~~矢倉~~
~~横歩取り~~
~~角換わり~~

~~横歩取り~~
~~角換わり~~
相掛かり

Alpha Zero
は相掛かりを有力視

個人的に面白かった点



~~7六歩~~　　2六歩

~~振り飛車~~
~~矢倉~~
~~横歩取り~~
~~角換わり~~

~~横歩取り~~
~~角換わり~~
相掛かり

Alpha Zero
は相掛かりを有力視

終盤が異常に辛い

普通に勝てばいいのに・・・

まとめ

- ボードゲーム AI には、探索部と評価部がある
  - Bonanza(全幅探索＋駒の相関)
  - Crazy Stone(選択探索＋モンテカルロ)
- AlphaGo は Deep leaning を上手く用いている
  - Policy Network (次の一手予想)
  - Value Network (勝率予想)
  - Rollout Policy (高精度モンテカルロ)
- Alpha Zero という汎用的な”手法”
  - 人間の知識は不要になった
  - チェス、将棋、囲碁をマスター
- これからのゲームAI
  - 不完全情報ゲームへの挑戦？


人狼知能
&
Poker