

コンピュータ囲碁

(アルファ碁について)

Sprint25
2019/1/5 (土)
友正 常雄



もくじ

- ・囲碁について
- ・直近1年間でどのような出来事があったか
- ・その分野に加わるために必要な知識や技術
 - ・ディープラーニング
 - ・強化学習
 - ・探索
 - ・アルファ碁の概要
- ・独自の考察

囲碁について



囲碁について

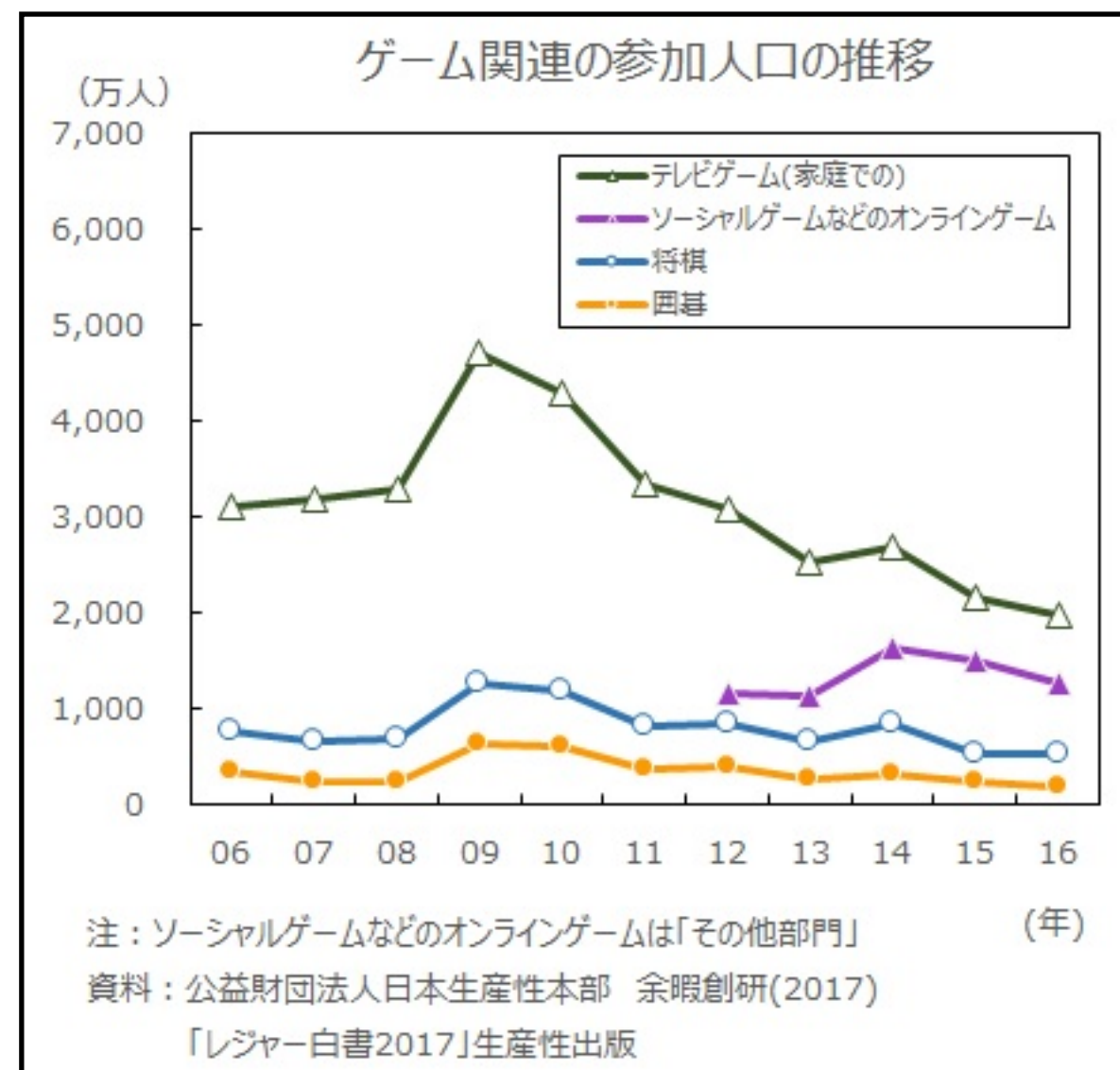
【概要】

囲碁（いご）とは、2人で行うボードゲームの一種。交互に盤上に石を置いていき、自分の石で囲んだ領域の広さを争う。単に碁（ご）とも呼ばれる。

【ルール】

2人のプレイヤーが、碁石と呼ばれる白黒の石を、通常19×19の格子が描かれた碁盤と呼ばれる板へ交互に配置する。一度置かれた石は、相手の石に全周を取り囲まれない限り、取り除いたり移動することはできない。ゲームの目的は、自分の色の石によって盤面のより広い領域（地）を確保する（囲う）ことである。

【囲碁人口】

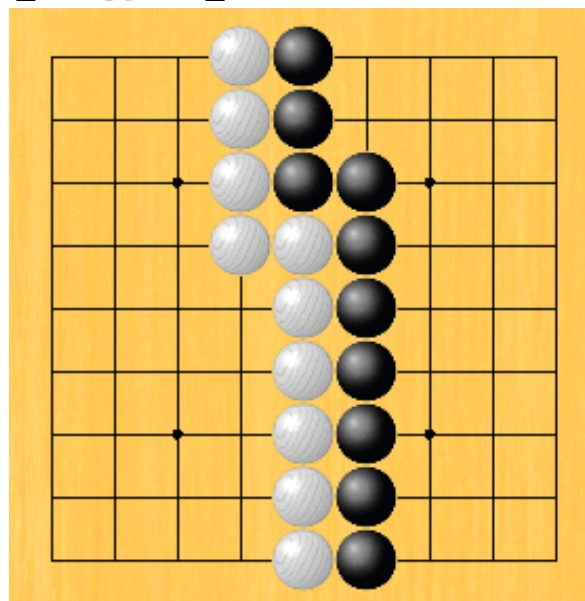


囲碁の世界競技人口は約3,600万人もあり、中国2,000万人、韓国900万人に次いで、日本は3番目となっている（日本棋院ホームページ『世界の推定囲碁人口』による）。

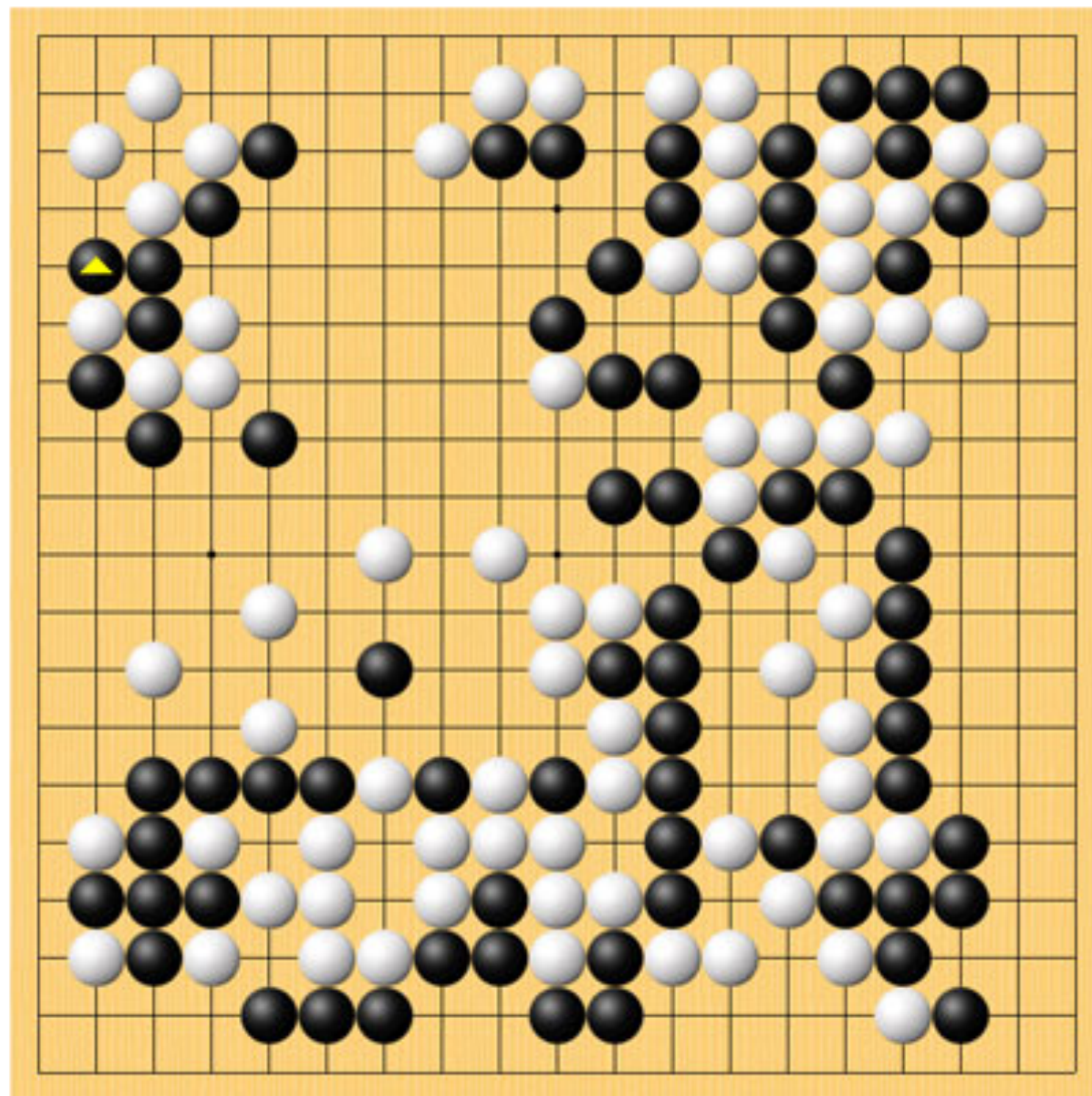


囲碁について

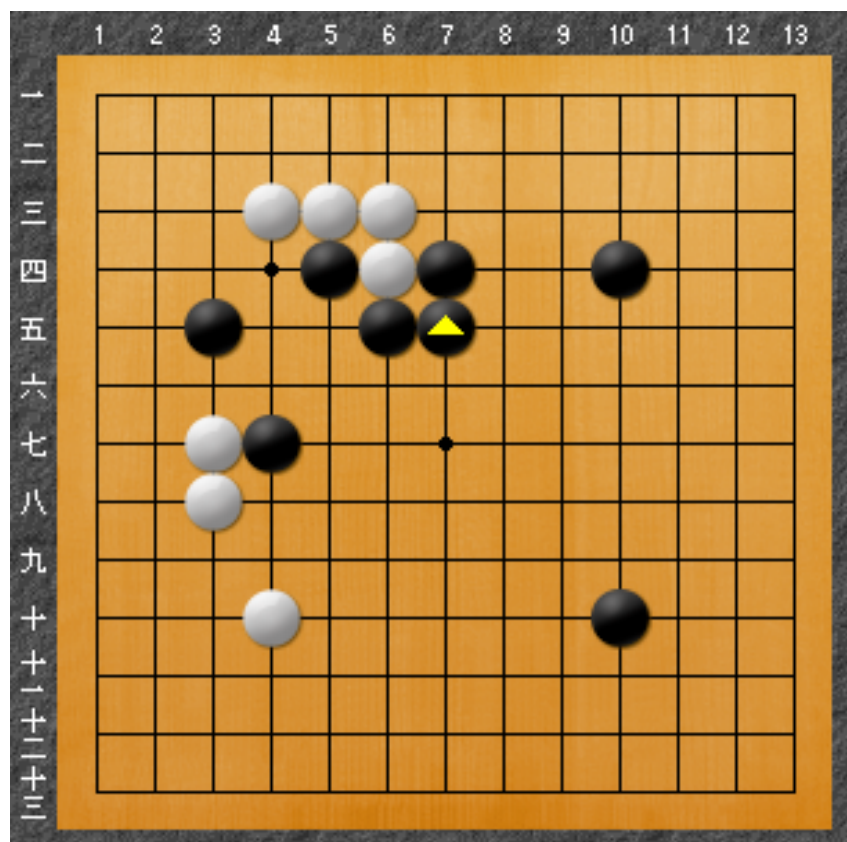
【9路盤】



【19路盤】



【13路盤】



直近1年間でどのような出来事があったか



直近1年間でどのような出来事があったか

2015年

- ・ 10月 アルファ碁が互先で初めてプロ棋士に勝利（初代）

2016年

- ・ 1月 DeepMind社がアルファ碁の論文発表
- ・ 3月 アルファ碁が韓国の李世ドル9段に勝利（2代目）

2017年

- ・ 5月 アルファ碁が中国の柯潔9段に勝利（3代目）
- ・ 10月 DeepMind社がアルファ碁ゼロ（4代目）

従来のバージョンとは大きく異なり、棋譜やビッグデータを必要とせず自己対局によって強化

2018年

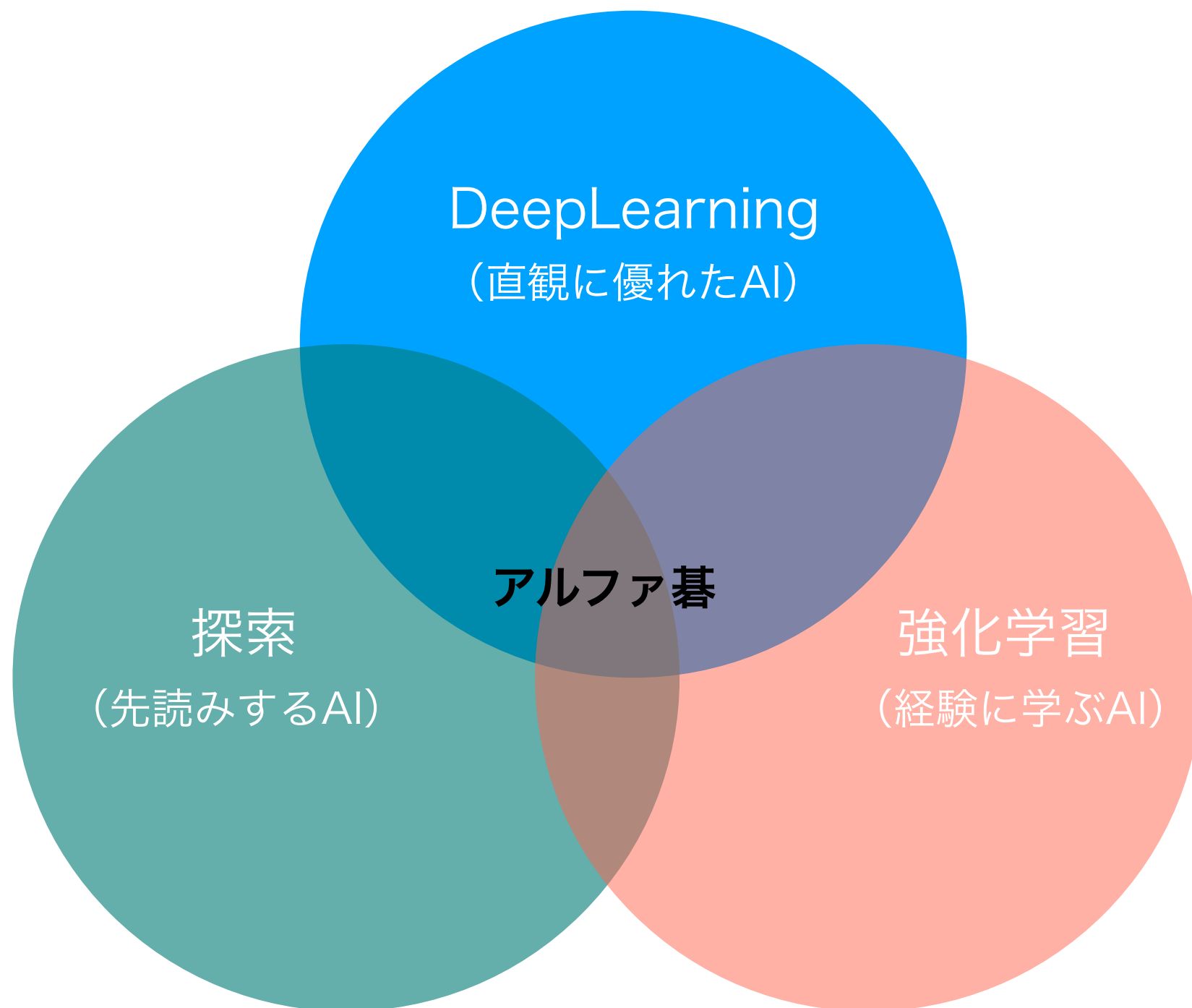
- ・ 8月 World AI Go Openが中国で開催
- ・ 12月 AlphaZero（5代目）

AlphaGo Zeroのアプローチを汎用化させ、2時間で将棋、4時間でチェスの最高峰のAIに勝利し、AlphaGo Zeroも8時間で上回った

その分野に加わるため
に必要な知識や技術



アルファ碁の主な仕組み

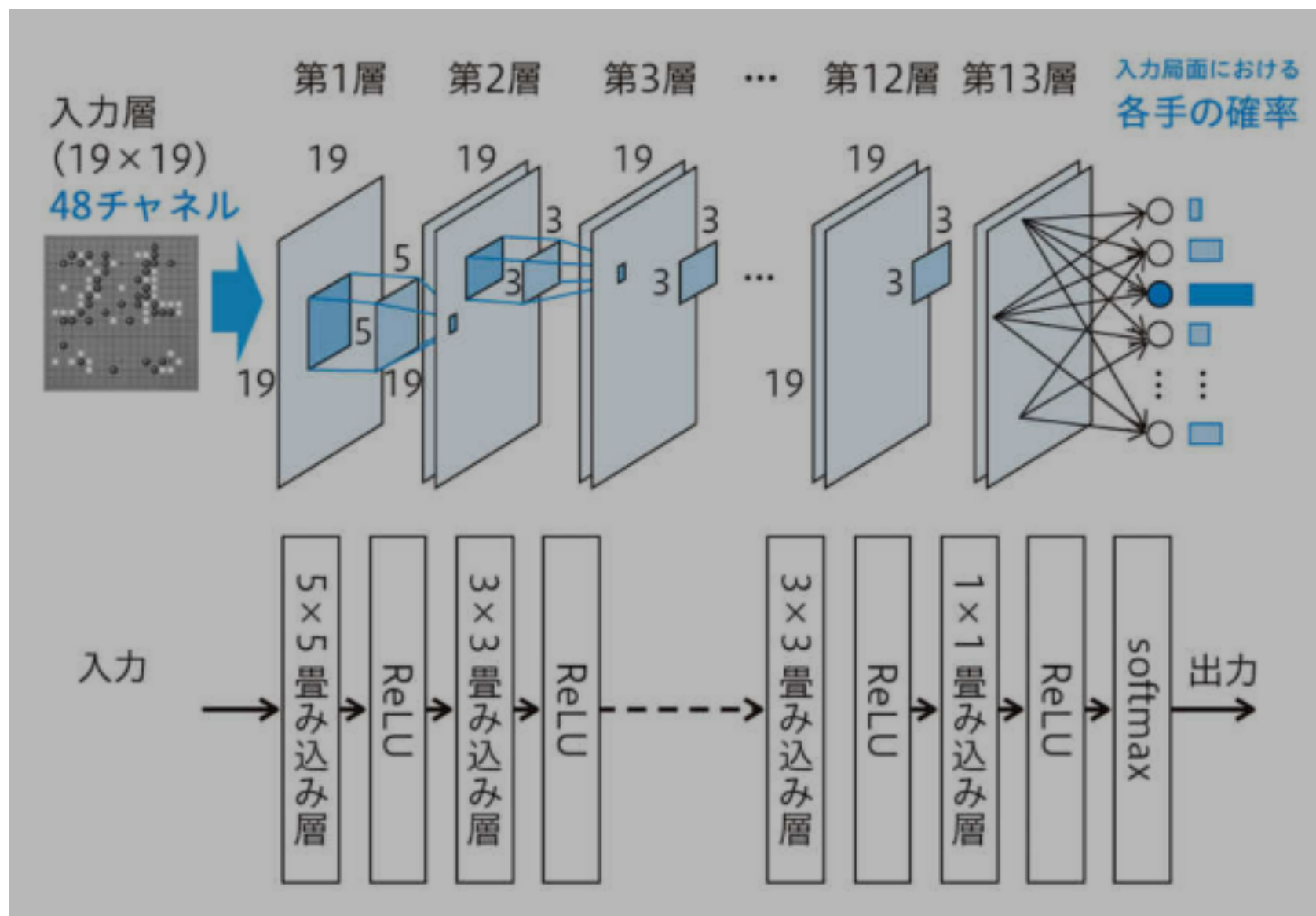




ディープラーニング

■ 囲碁の手を選択するCNN (SLポリシーネットワーク)

- ・局面を入力して、各位置に打つ確率の予測値を出力する
 - ー入力は48チャンネル
 - ー全部で13層
- ・1～12層のフィルタは各層に192種類ずつあり、1層目のみ5×5、2～12層は3×3





ディープラーニング

■囲碁の手を選択するCNN(SLポリシーネットワーク)

・入力層の48チャンネルの特徴

入力チャンネルの種類	チャンネル数
黒石の位置	1
白石の位置	1
空白の位置	1
k 手前に打たれた位置 ($k=1\sim 8$)	8
石がある場合の当該連の呼吸点の数 ($k=1\sim 8$)	8
そこに打った後、石を取れるか (取る数: $k=1\sim 8$)	8
そこに打った後、当該連を取られる場合に、何個石を取られるか? (石の数: $k=1\sim 8$)	8
そこに打った後の、当該連の呼吸点の数 (呼吸点の数: $k=1\sim 8$)	8
そこに打った後、隣接する相手の連をシチョウで取れるか?	1
そこに打たれた後、隣接する味方の連をシチョウで取られるか?	1
合法手か?	1
すべて1で埋める	1
すべて0で埋める	1
合計	48



ディープラーニング

■囲碁の手を選択するCNN(SLポリシーネットワーク)

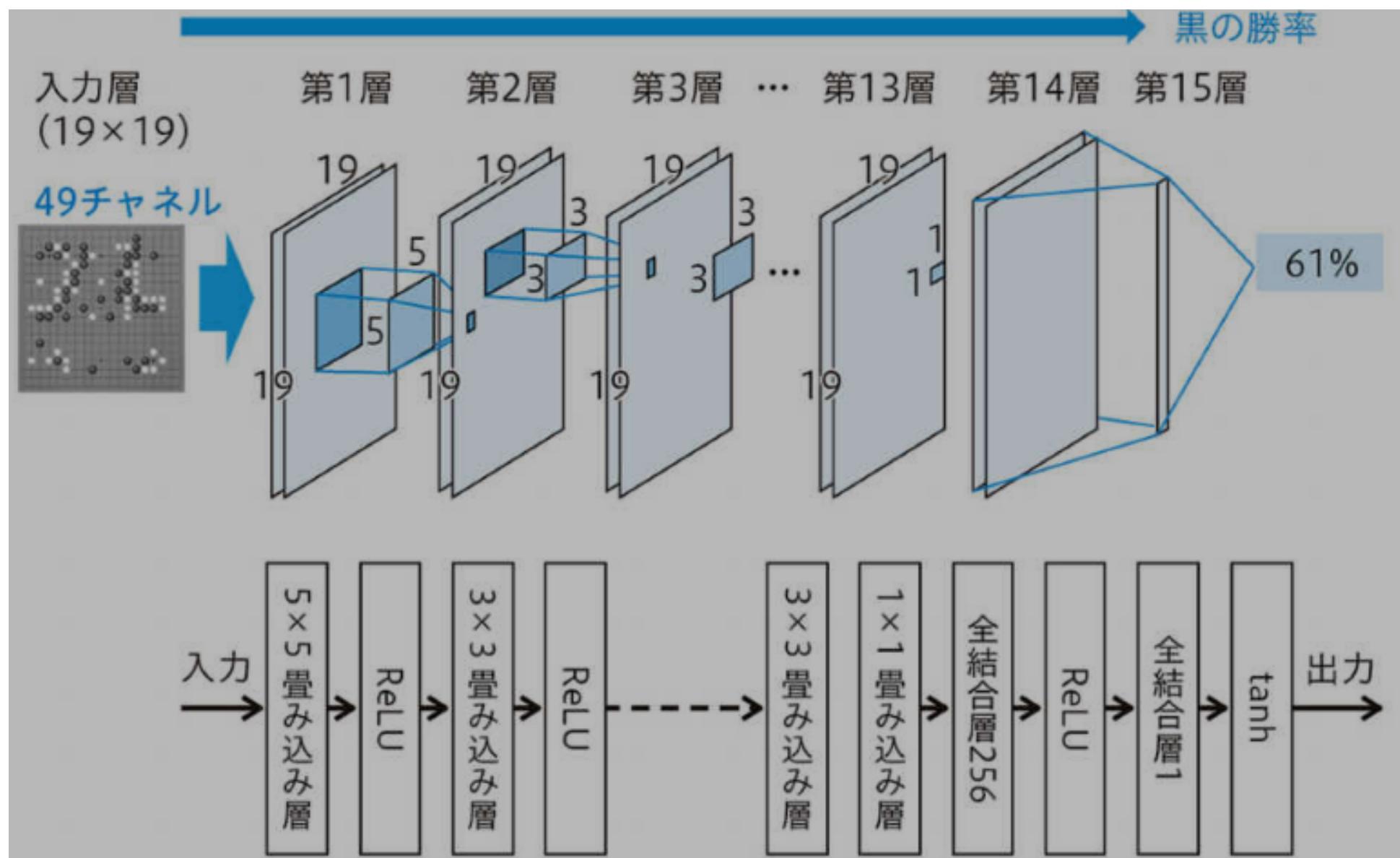
- ・SLポリシーネットワークは400万個もあるフィルタ重みパラメータを学習するには、大量の高品質な学習データ(入力局面と正解ラベルの組み)が必要です。
- ・高品質とは、この場合強いプレイヤーの手が大量に必要となります。そこで、KGS(会員数10万人を超え)というインターネット対局サイトの6段以上のプレイヤーによる対戦棋譜16万局分を使いました。
- ・1局の棋譜は200手程度の手順からなるので、1局から200個程度の学習データが得られます。結果、この棋譜から合計約3,000万個($\asymp 16\text{万局} \times 200\text{個}$)の学習データを作ることができます。
- ・さらに、回転・反転させることで8倍の学習データに拡張しています。
- ・結果として、 $3,000\text{万} \times 8 = 2\text{億}4,000\text{万}$ の学習データを使っていることに相当します。



ディープラーニング

■局面の有利不利を予測するCNN(バリューネットワーク)

- ・バリューネットワークは、局面を入力とし、勝率予測値を出力するCNN
- ・バリューネットワークの構成
 - ー入力は49チャンネル(SLポリシーネットワークと同じ48チャンネル+手番)
 - ー全部で15層
- ・1～12層のフィルタは各層に192種類ずつあり、1層目のみ5×5、2～12層は3×3





強化学習

■アルファ碁の強化学習は、
SL (Supervised Learning) ポリシーネットワークを、より多く勝てるRL (Reinforcement Learning) ポリシーネットワークに「強化」するため、
強化学習を利用しています。

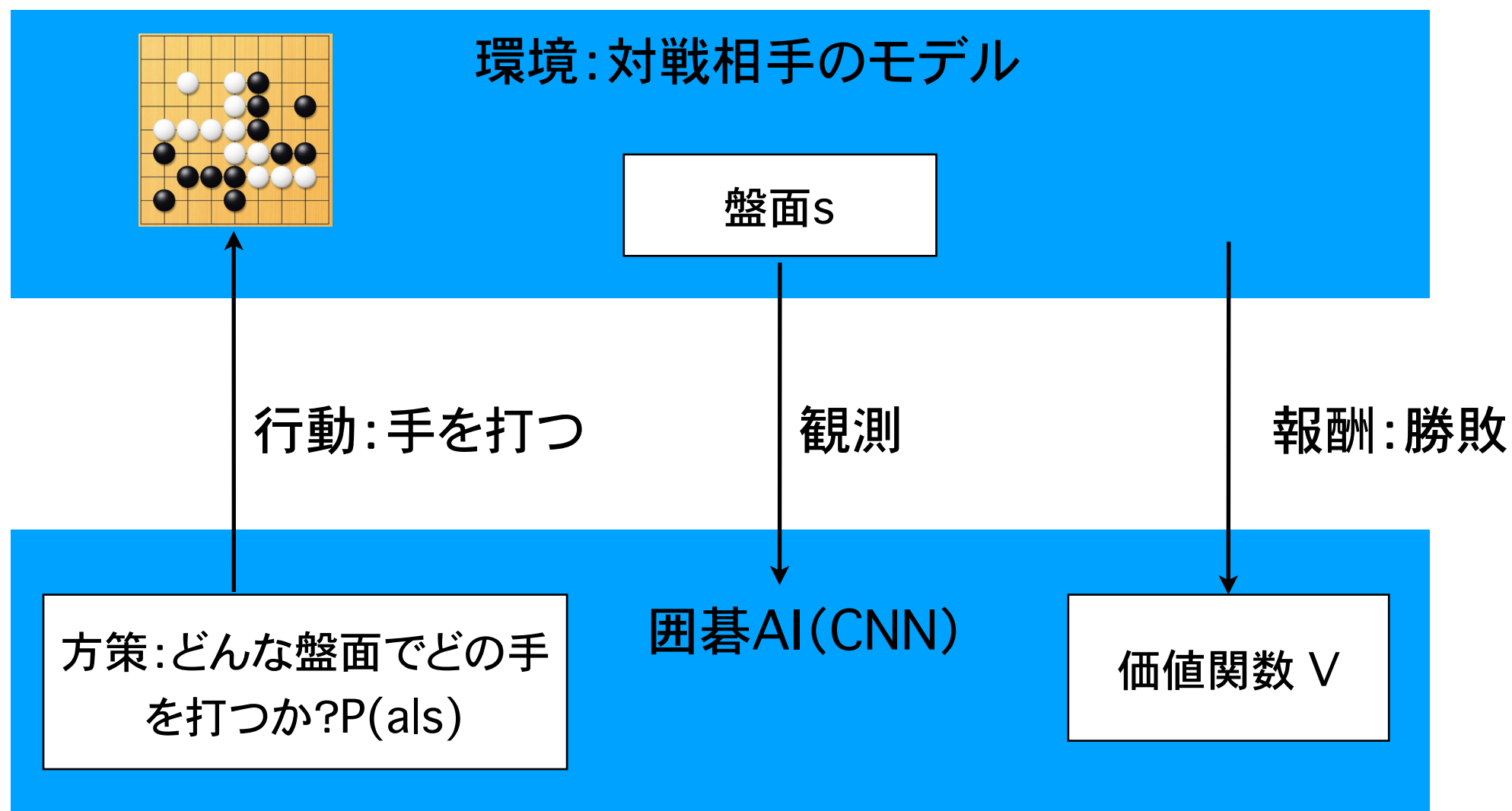
【RLポリシーネットワークを獲得するための強化学習の概要】

- ・学習手法: SLポリシーネットワークを初期値とし、「ゲームの勝利」を報酬として、方策勾配法により強化学習
 - 一勝った時は、勝に至る手をできるだけ選ぶようにパラメータを更新
 - 一負けた時は、負けるに至る手をできるだけ避けるようにパラメータを更新



強化学習

■アルファ碁の強化学習は、
SL (Supervised Learning) ポリシーネットワークを、より多く勝てるRL (Reinforcement Learning) ポリシーネットワークに「強化」するため、
強化学習を利用しています。



※オリジナルのSLとの直接対決では、80%勝てるようになりました。



探索

■2人ゼロ和有限確定完全情報ゲームでは
「先手必勝」「後手必勝」「引き分け」のいずれかを「数学的には」確定できることが知られています。

- ・○×ゲーム : 引き分け
- ・5目並べ : 先手必勝
- ・どうぶつ将棋 : 後手必勝
- ・6×6のオセロ : 後手必勝
- ・8×8のオセロ : 不明
- ・将棋／囲碁 : 不明

■一方、囲碁、将棋、チェスのように、局面の分岐・組み合わせの数が非常に大きい場合、完全解析は事実上不可能です。

勝ち負けが数学的に決定できることと、その解が簡単に得られることは全く意味が異なるのです。



探索

■SLポリシーネットワーク

SLポリシーネットワークでは、1手の先読みをしていると考えられます。この1手を、2手、3手、…、 d 手と増やして評価の精度を上げていくこう、というのがゲームにおける「探索」の考え方です。

■ただし、深さ d 手として、その局面の合法手の数が w 個あるとすると、単純には w の d 乗個の局面を評価しなければならず、計算量が爆発します。

■これは指数オーダと言われ、 d が大きくなるとすぐに膨大な数となってしまいます。したがって、探索の問題では、この膨大な探索空間をいかに効率的に探索するかが焦点となります。



探索

■モンテカルロ木探索

「有望な手」をより深く調べることで、探索の精度を高める

「モンテカルロ木探索」という手法です。以下のStep1～4のシミュレーションを制限時間一杯まで繰り返し、最終的に最も試行回数の多いノードを選択する。

Step1：子ノード選択（Selection）

リーフノードに至るまで、UCB1（=バイアス+勝率）を最大化する子ノードを選択して手を進め、木を降りていきます。

Step2：展開（Expansion）

訪問回数がNthr以上となった場合に、新しいノードを作成します。

Step3：評価（Evaluation）

プレイアウトを実行します。

Step4：更新（Backup）

プレイアウトが終わった後は、ルートノードに至るすべてのノードについて、プレイアウトの勝敗をゲーム木に反映します。

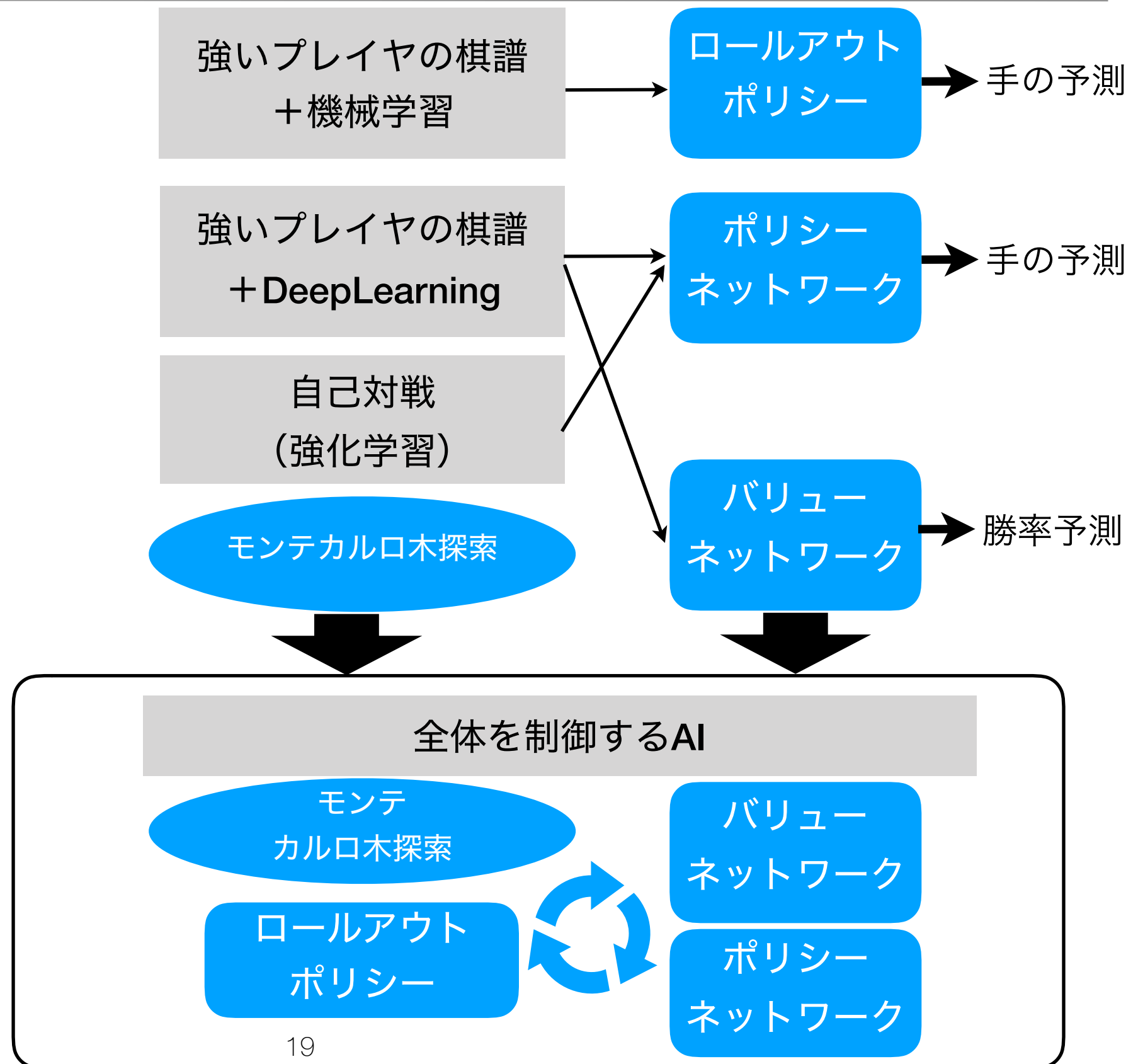


アルファ碁の概要

【3つのポリシー】

- ・ ロールアウトポリシー
- ・ ポリシーネットワーク
- ・ バリューネットワーク

【モンテカルロ木探索】





考察

【全産業への展開】

・ゲームの中でも難しいとされているが囲碁が攻略されたので、近い将来、まだコンピュータでは無理といわれているさまざまな領域にDeepLearning・強化学習などの技術が応用されて浸食してくるのは時間の問題だと思います。

【囲碁界】

・トッププロが敗れた時は、囲碁界は騒然としましたが、今は更に強くなるためにAIを活用しているので囲碁界のレベルアップにつながっている。これは囲碁界に限らずに、どの分野でも同じようなことが起こると思います。

・人間同士の対局に価値が無くなると考えていましたが、決してそうでは無く、人間の能力を拡張するための道具として利用している。

・2018年8月にWorld AI Go Openも開催されて、コンピュータ囲碁も今後も盛り上がっていくと思います。



Appendix

■2人ゼロ和有限確定完全情報ゲーム

(ふたり ぜろわ ゆうげん かくてい かんぜんじょうほう ゲーム)

ゲーム理論によるゲームの分類のひとつで

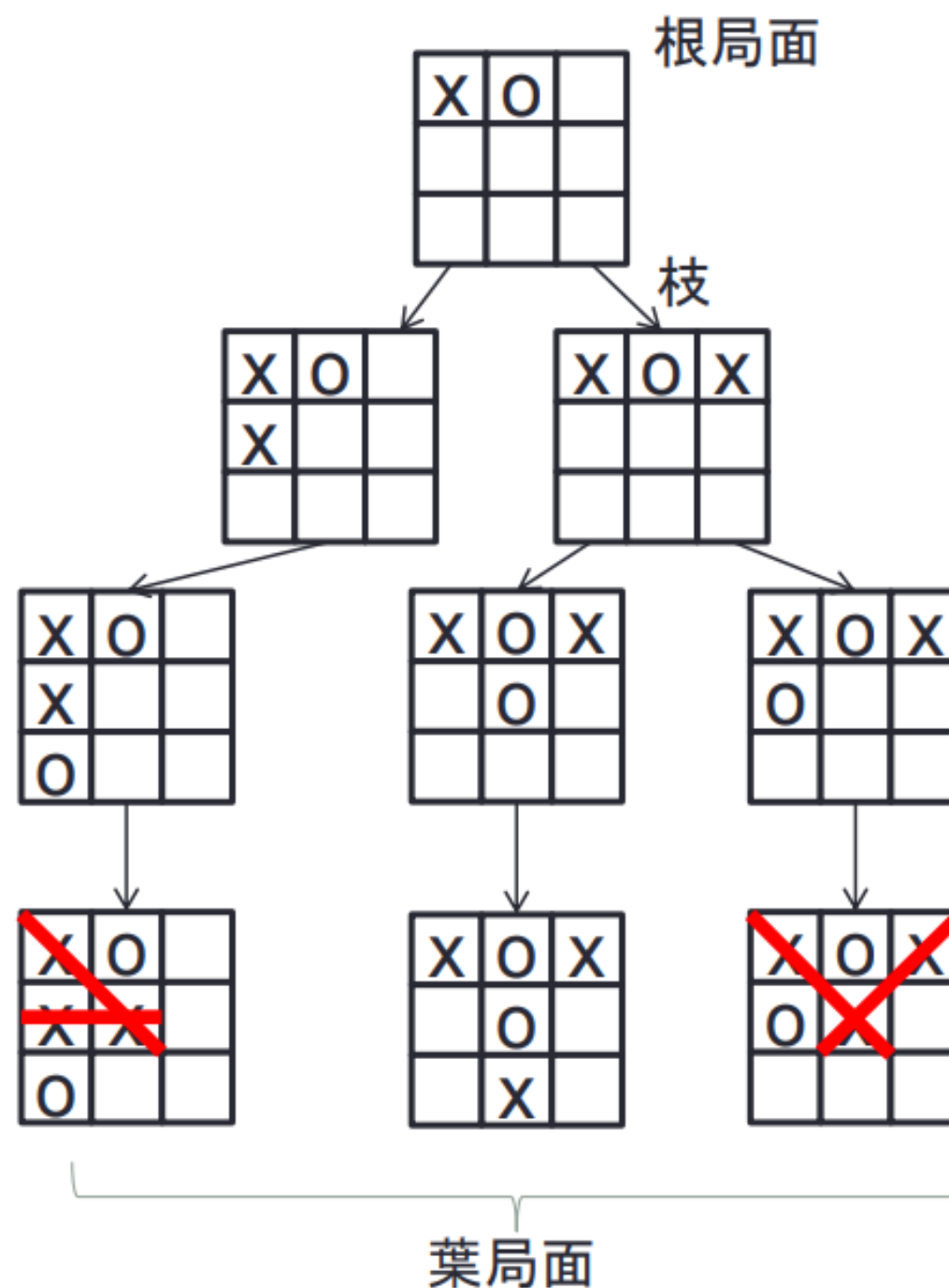
- ・二人:プレイヤーの数が二人
- ・ゼロ和:プレイヤー間の利害が完全に対立し、一方のプレイヤーが利得を得ると、
それと同量の損害が他方のプレイヤーに降りかかる
- ・有限:ゲームが必ず有限の手番で終了する
- ・確定:サイコロのようなランダムな要素が存在しない
- ・完全情報:全ての情報が両方のプレイヤーに公開されている



Appendix

■ゲーム木

- ゲームを木として表す
 - 局面: 節点
 - 手: 枝
 - 初期局面: 根
 - 末端局面: 葉
- 全て展開すれば解ける!
 - 勝ち負け引分が分かる





Appendix

■合法手

ある局面でルールを満たす候補手のこと。

ゲーム木を考える場合、合法手の数は枝分かれの数になります。

この枝分かれの数は、分岐因子と呼ばれることもあり、ゲームの難しさの指標の一つになっています。

■ロールアウトポリシー

アルファ碁において、手の表面的な特徴を元に、その手が打たれる予測確率を出力する数理モデルの1つです。高速に評価できる特徴のみを利用するため、評価速度は速いですが、手の予測確率はやや落ち24%程度に留まります。

■プレイアウト

ある局面から、黒番と白番が、短い試行時間で交互に手を選択し、終局（勝ち負けが判断できる局面）まで進める手法です。

このプレイアウトを用意に実行できることが、原子モンテカルロやモンテカルロ木探索を行う条件となります。

確率ベースのシミュレーション手法です。



Appendix

■参考文献

書名:最強囲碁AI アルファ碁 解体新書

著者:大槻 知史

監修:三宅 陽一郎

発行:翔泳社

■参考URL

書名:https://www.slideshare.net/shogotakeuchi/ss-62415546?from_action=save

作者:北海道大学研究員 竹内 聖悟