



研究背景と目的

強化学習手法の台頭

- DQN, A3C (Atari 2600)
- AlphaGo Zero (囲碁), Alpha Zero (将棋・チェス)

人間の知識なしで人間を超えた

囲碁将棋を主な対象に大規模強化学習の知見を蓄積

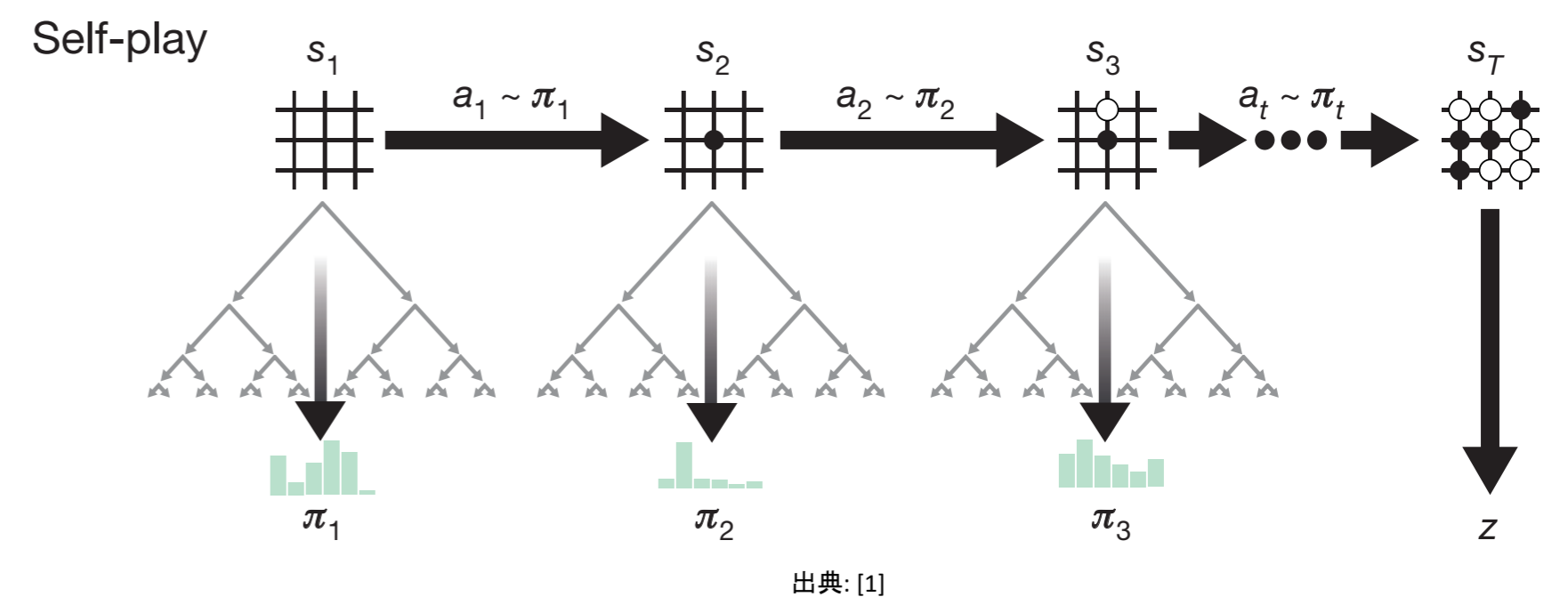
- 探索の効率化による学習の高速化
- 計算資源の削減

学習結果の可視化と説明

- 「なぜこの局面だとこの手が評価が高いのか」

一方で、

- 計算資源の肥大化
AlphaGo Zero の再現には 1700 年かかる
- 学習結果の信頼性の検証
なぜこの結果になるのか人間に説明
という課題



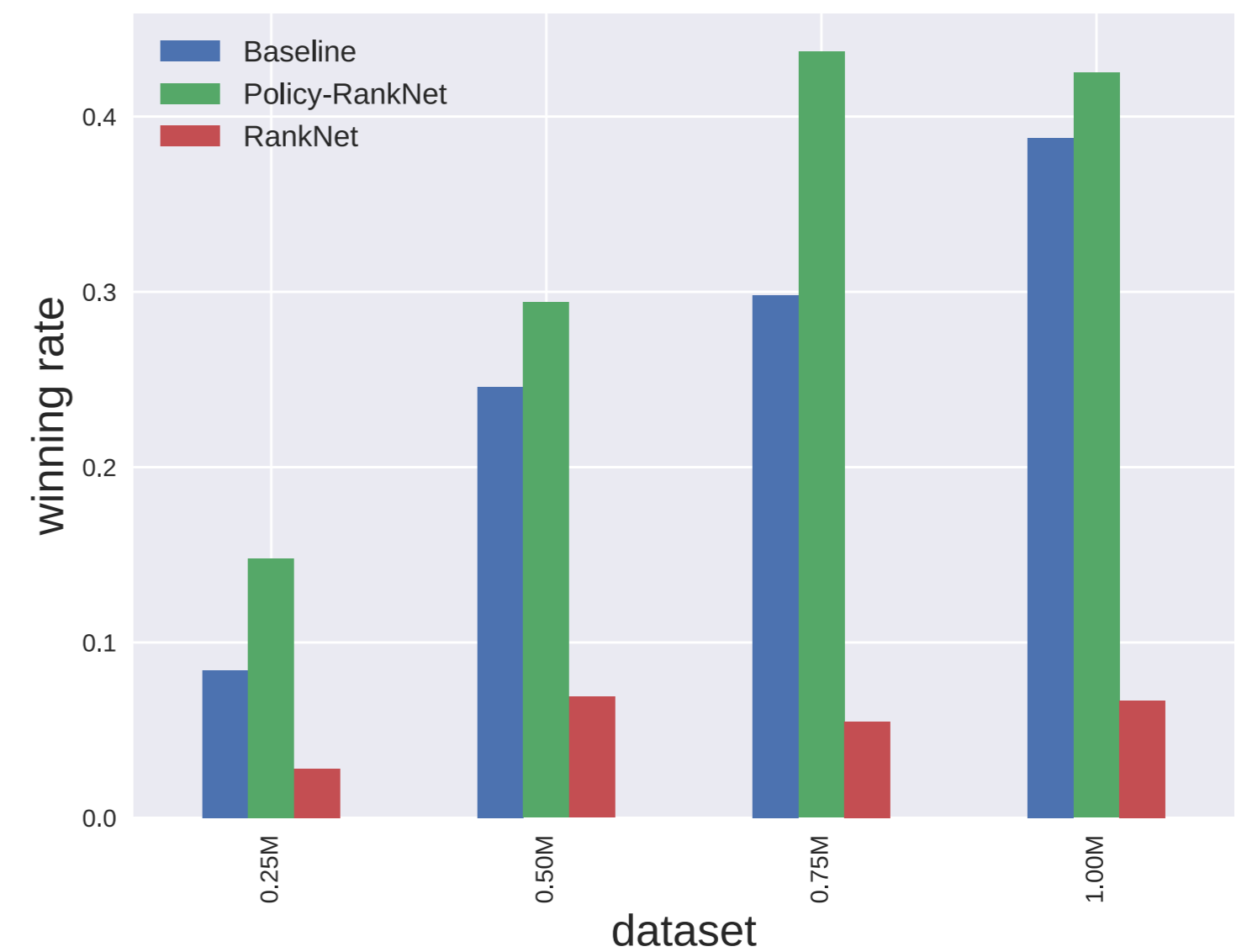
出典: [1]

局面对による学習の高速化

2つの局面の優劣を学習することにより必要な学習局面 (i.e. 自己対戦回数) を削減

→ Learning to Rank によって価値関数を学習
学習のデータ数を二乗のオーダーで増やせる

既存手法と比較して全体的に性能が向上

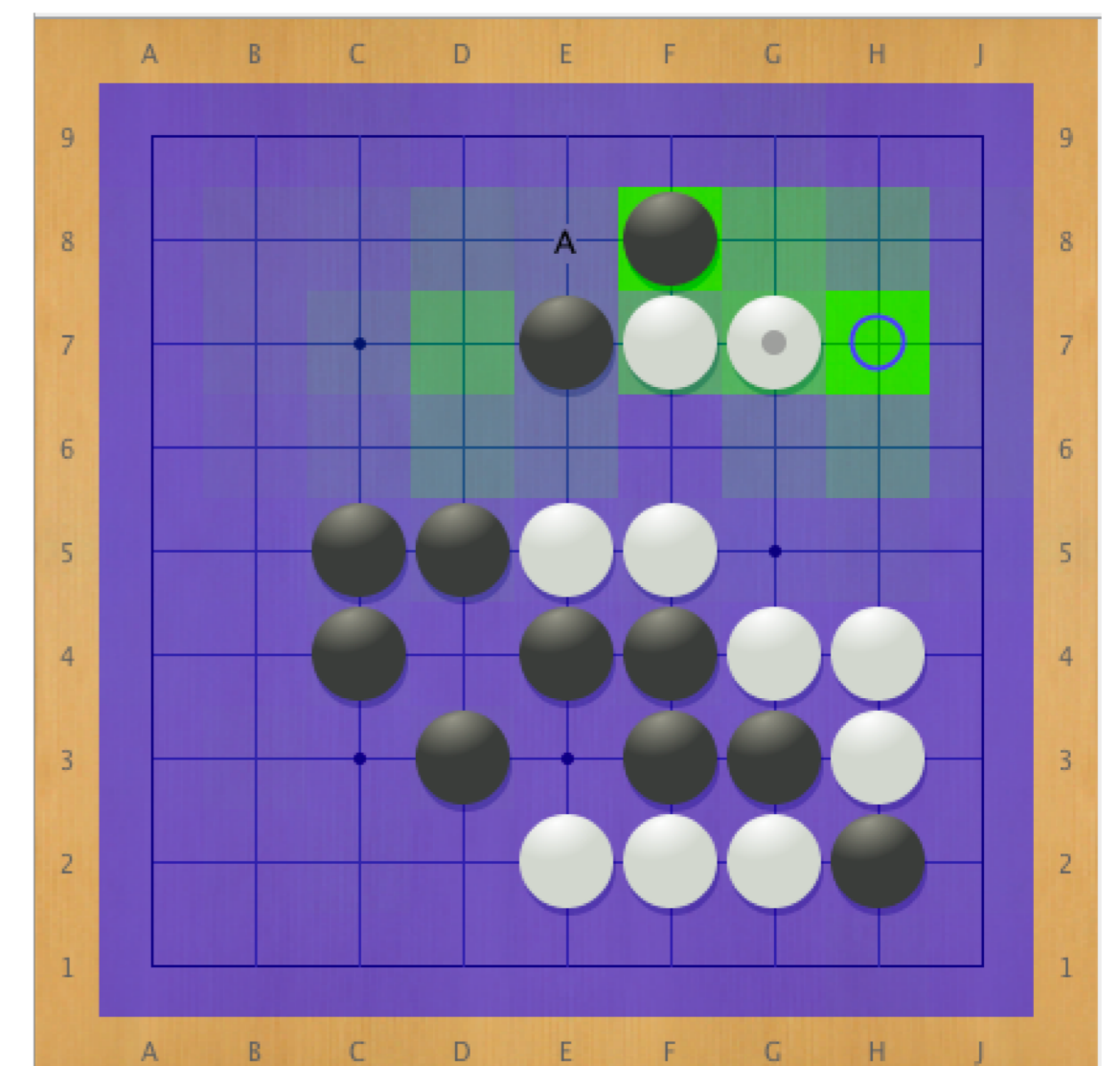


学習器の説明

学習したニューラルネットワークが「よく見ている」場所を可視化
Saliency Map を用いて最も注目している位置をハイライト

今後

- 探索部分の効率化を行い、学習を更に高速化
 - まだ良くわかっていない局面をより重視するような探索
- 探索を組み合わせた学習器の説明



References

1. D. Silver et al, "Mastering the game of Go without human knowledge", In: Nature 24270
2. O. David et al, "DeepChess: End-to-End Deep Neural Network for Automatic Learning in Chess", In: ICANN 2016