

# 対戦型パズルゲームにおける 機械学習AIと人間の知識を用いたAI の比較

電気通信大学 情報理工学部 卒業論文発表

総合情報学科 メディア情報学コース

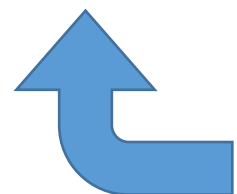
橋山研究室

1310163

柴澤弘樹

# 背景：ゲームAIの現状

- AIがプロの人間プレイヤーに勝つ
  - AlphaGo……囲碁
  - DQN ……Atari 2600ゲームの29種類



機械学習（ディープラーニング）の発展

## 問題

- 内部処理がブラックボックス
- 処理、挙動の解釈が難しい
- 学習結果の再利用が難しい

# 目的

- 機械学習AI
  - 強い、事前知識不要
  - × 学習結果の解釈が難しい
- ルールベースAI
  - 処理の解釈、改良が容易
  - × 知識のルール化が難しい

2つの手法を比較、検討

# ぷよぷよ

- 色付きブロック「ぷよ」が降下
- 同色ぷよを4つ繋げると消える
- 連続で消すと連鎖
- ぷよの予告は3手先まで

- 連鎖=>スコア=>おじゃま
- 大連鎖が勝利へ
- 難しさ：先読み、長期的視点の必要性



# 連鎖ポテンシャル法<sup>[2][3]</sup>

- 見えている手のみ（3手分）を全探索
- 1手目での消去なし
- 2, 3手目で発動する連鎖のスコアを最大化



改良

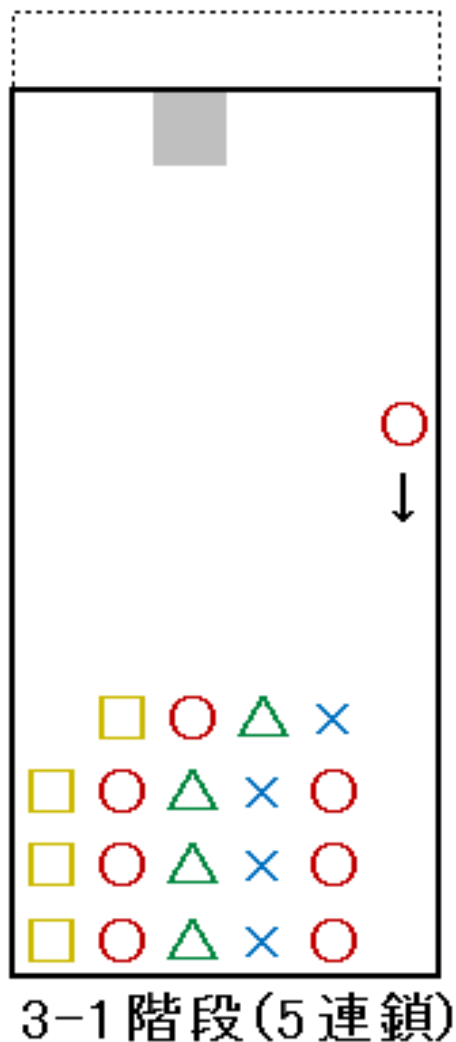
3手目を全幅探索 + 消去の許可

---

[2] 富沢大介, 池田心, シモンビエノ. 落下型パズルゲームの定石 形配置法とぷよぷよへの適用. 情報処理学会論文誌, Vol. 53, No. 11, pp. 2560–2570, nov 2012.

[3] 大月龍, 前田新一, 石井信. 不完全情報ゲームに対する階層化したモンテカルロ探索とそのぷよぷよへの適用. 電子情報通信学会技術研究報告. NC, ニューロコンピューティング, Vol. 113, No. 500, pp. 275–280, mar 2014.

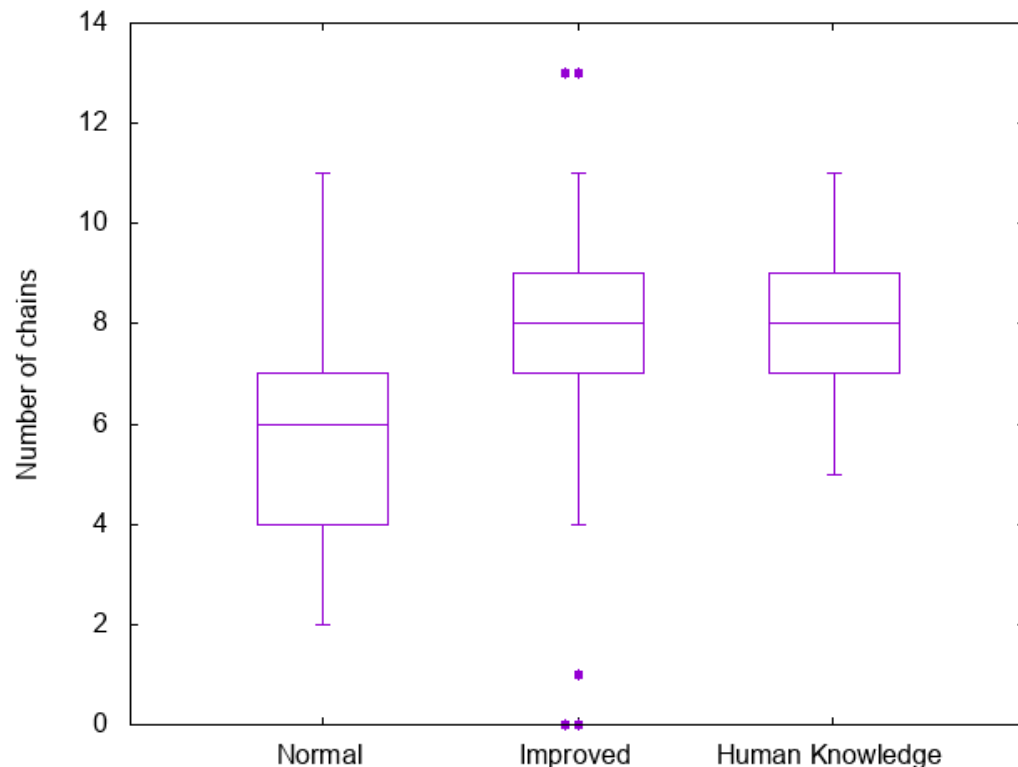
# 人の知識を適用したAI



- 知識
  - 3-1階段の構築ルール
  - if-thenルールを書き下し（設置法20種）
- 実装
  - 初手6手に適用＋連鎖ポテンシャル法
  - 連鎖数の安定を目指す

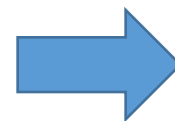
# 実験1：連鎖構築シミュレーション

- ポテンシャル法、人の知識を適用したAI
- 実験条件：32手+発動1手、50試行（同配石）



平均連鎖数

ポテンシャル最大化法	5.96
ポテンシャル法の改良法	7.78
人の知識を適用したAI	8.10



人の知識で連鎖数が安定

# 実験2：対戦

- ゲーム内のAIと対戦
  - DQN：学習環境と同条件で対戦
  - 人の知識適用AI：スコア閾値2100点で連鎖発動

勝利数

	実装AI - ゲーム内AI
DQN	2 - 48
人の知識適用AI	24 - 26

平均スコア

	実装AI	ゲーム内AI
DQN	743.30	1790.12
人の知識適用AI	4762.52	2703.74

人の知識を適用したAIが優れていた



# 考察

- 機械学習よりルールベースAIが強い
  - 機械学習は解釈・改善が難しい
  - 知識のルール化は困難な場合あり
- 

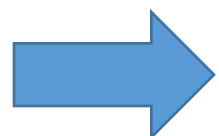
# 今後の課題

- 人の棋譜を機械学習
- 機械学習と知識抽出の組み合わせ
- 人を楽しませるAIへ

# まとめ

- 機械学習AIと人の知識を適用したAIの比較
- 対戦型パズルゲーム「ぷよぷよ」を対象
- 人の知識を適用したAIが強かった

- 機械学習：解釈が困難
- 知識適用：ルール化が困難



今後2つの手法を組み合わせることを検討

# DQN: Deep Q-Learning<sup>[1]</sup>

- Q値をディープラーニングで学習
- パラメータ $\theta_t$ の更新式

$$\theta_{t+1} = \theta_t + \alpha \left( R_{t+1} + \gamma \max_a Q(s_{t+1}, a; \theta_t) - Q(s_t, a_t; \theta_t) \right) \nabla_{\theta_t} Q(s_t, a_t; \theta_t)$$

$t$ :時刻,  $\alpha$ :学習係数,  $R_{t+1}$ :  $t \rightarrow t+1$ で得た報酬,  $\gamma$ : 割引係数,  
 $s_t$ :  $t$ での状態,  $a_t$ :  $t$ での行動

- ゲーム内AIとの対戦で戦術を学習
- 報酬 $R_{t+1}$ : (自スコア-相手スコア)の変化

---

[1] Volodymyr Mnih et al. Human-level control through deep reinforcement learning.  
*Nature*, Vol. 518, No. 7540, pp. 529–533, 02 2015.