# 深層強化学習を用いた麻雀 AI に関する研究

大神卓也[1]・天野克敏[2]・鶴岡慶雅[1]

[1] 東京大学大学院情報理工学系研究科 [2] 東京大学大学院学際情報学府

## 研究の目的

近年，深層強化学習を用いた AI が囲碁やポーカーなどのゲームでプロのプレイヤに勝利.

麻雀でも強力な AI が研究されているが，プロのプレイヤに勝利した例はない.

本研究では，AI の評価が行われるネット麻雀とプロの麻雀のドメインの差に起因する課題を解決することを目指す

## 異なる順位点における打牌選択

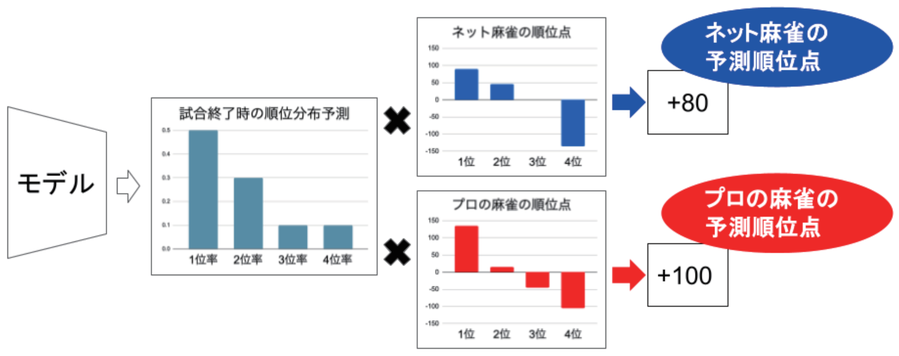
### 課題

麻雀ではルールによって，試合終了時の順位に応じてもらえる順位点が異なる．既存の麻雀 AI はネット麻雀に特化しており，プロのルールに適応していない.

### 解決策

強化学習のエピソード内の時点から，試合終了時の順位分布を予測するモデルを構築.

予測順位分布と各ルールの順位点の積から予測順位点を算出.

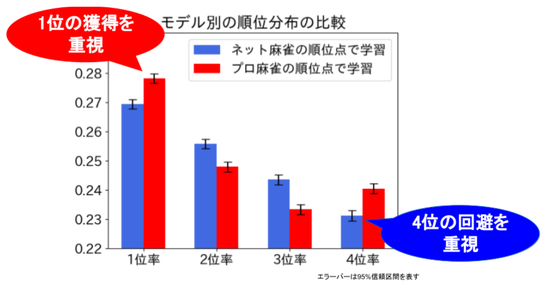



### 結果

ネット麻雀の順位点で学習したモデルは４位率の低下

プロの麻雀の順位点で学習したモデルの１位率の向上

→ 順位点の違いによる選択の違いの傾向が見られた




## Transformer を用いた手牌推定

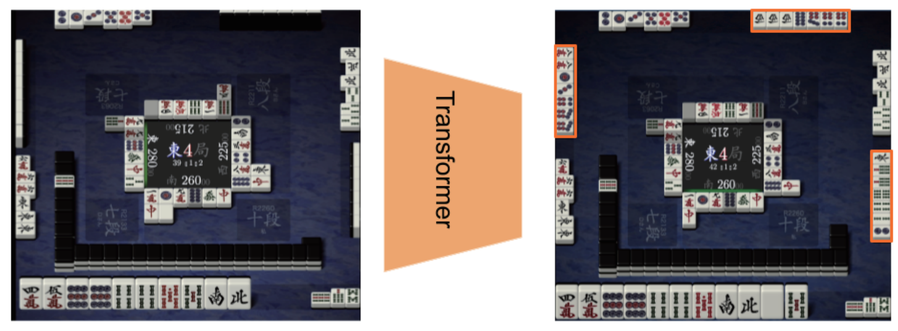
### 課題

麻雀の上級者は相手の手牌を推定し，他のプレイヤーよりも有利に戦う．一方，既存の麻雀 AI はこのような相手の手牌の推定が苦手であることが指摘されている.

### 解決策

対戦相手の手牌を系列として推定するために Transformer を用いた教師あり学習を利用.

プロのフィードバックをもとにした評価指標の策定と適切なデコード手法の選定による性能向上.




### 結果

提案手法（ビームサーチ，サンプリング）において，ベースラインよりも高い一致率と低いシャンテン数誤差を記録.

| | 一致率 | シャンテン数誤差 | ヒストグラム交差 | ペア一致率 |
|---|---|---|---|---|
| ランダム選択 | 0.302 | 1.586 | **0.924** | **0.309** |
| CRF | 0.389 | 1.083 | 0.774 | 0.748 |
| ビームサーチ（幅 10） | **0.463** | 0.717 | 0.851 | 0.840 |
| サンプリング ($p = 0.7$) | 0.426 | **0.690** | 0.921 | 0.421 |

## 試合の評価における分散の削減

### 課題

麻雀のような不完全情報ゲームでは，プレイヤの実力以外の要素（運）によって結果が左右される.

そのため，試合結果からプレイヤの実力を評価するためには膨大な試合数を必要とする.

| 試合 | 1 | 2 | … | n | 平均 | 分散 |
|---|---|---|---|---|---|---|
| 平均順位 | 1 | 4 | … | 2 | 2.5 | 1.25 |
| 推定値 | 2.6 | 2.4 | … | 2.3 | 2.5 | 0.25 |

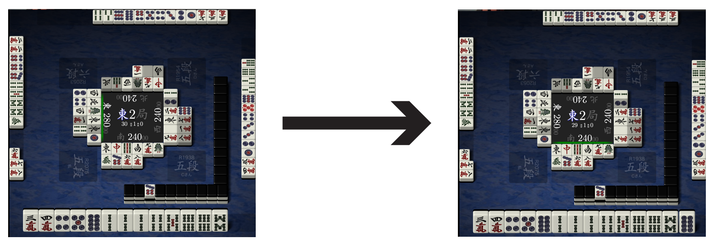
平均順位の 95% 信頼区間

$$2.5 \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

分散が小さければ，少ない試合数で同じ信頼区間が求まる

### 解決策

（運）= ( チャンスプレイヤの行動前後の価値関数の変動 )



V(h) → V(ha)

+1000 → +2500

利得の分散を最小化させる V を求める

$$\underset{Luck}{\textbf{Minimize}} : Var(Utility - Luck)$$

$$\begin{aligned} Utility &= (\text{サンプルの利得}) \\ Luck &= \sum V(ha) - V(ha) \end{aligned}$$