

# 「ゲシエンク」AIのためのニューラルネットワーク評価関数の設計

野本 瑛二 【高度プログラミング研究室】

## 1 はじめに

多人数・不完全情報ゲーム「ゲシエンク」は、プレイヤー間で情報差がなく、かつ、各局面でプレイヤーがとれる行動が高々2つであるというゲーム特徴を有する。ゲームルールこそシンプルであるが、現時点で明確な有効戦略は確立されていない。著者はこれまでに、ルールベースプレイヤー [1] および深層学習プレイヤー [2] を実装したが、後者は前者に上回るには至らず、やや劣る結果であった。本研究では、複数のニューラルネットワークモデルを実装し、対戦実験を行うことで、「ゲシエンク」におけるAIプレイヤーの評価関数として有効なニューラルネットワーク設計を検討する。

## 2 ネットワーク設計

本研究では、以下の3種類のニューラルネットワークを設計した。このとき、各ネットワークにおける学習パラメータ数が800,000程度となるように調整を行った。

**MLP** 全結合層5層からなるネットワーク。

**CNN** 畳み込み層1層と全結合層3層からなるネットワーク。

**ResNet** 畳み込み層5層（畳み込み層2層からなる残差ブロック2つを含む）と全結合層3層からなるネットワーク。

入力は、ある1局面における各行動後の局面情報とその行動をone-hot形式のバイナリテンソルとして表現したものであり、出力は、各行動の行動価値とした。

## 3 実験方法・実験結果

各ニューラルネットワーク評価関数を用いるプレイヤーを実装し、ルールベースプレイヤー [1] 3体との4人対戦を通じてそれぞれ学習を行った。共通条件として、バッチサイズは1024、最適化手法はAdam（学習率は0.0001、その他のパラメータはPyTorchの既定値）、損失関数は平均二乗誤差、学習時間は72時間、ターゲット値は $0.9^k v$ （ $k$ はゲーム終了までの対象プレイヤーの手数、 $v$ は対象プレイヤーの最終スコア）とした。学習における「対象プレイヤー」について、先行研究 [2] に倣い、ルールベースプレイヤーの行動を学習する教師あり学習から、自身の行動を学習する強化学習へと徐々にシフトする手法を採用した。学習開始から60時間以降は強化学習のみとなる。

学習過程における勝率推移を図1に示す。各モデルを6時間ごとに保存し、保存時点ごとにルールベースプレイヤー3体との10,000試合の勝率をプロットした。また、ネットワークによって学習速度が異なることから、累積学習局面数の推移を図2に示す。

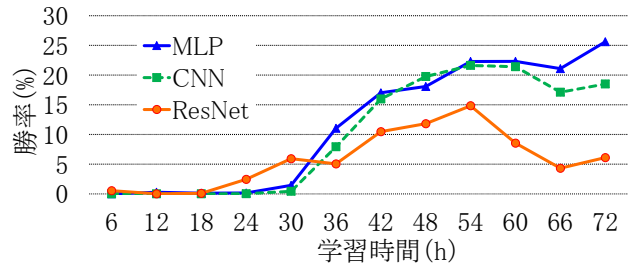


図1: 学習過程における勝率推移

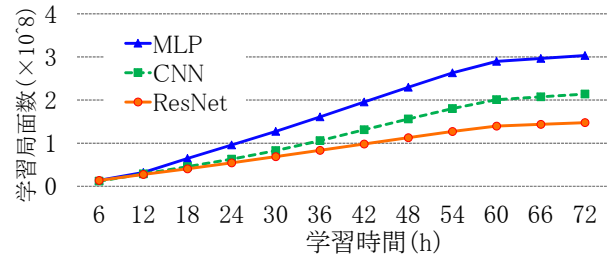


図2: 累積学習局面数の推移

学習時間60時間あたりで勝率の低下が見られた要因としては、教師あり学習+強化学習から完全な強化学習に移ったことによる学習対象の変化が考えられる。また、ResNetよりも簡潔に構成したCNNの方がよい成績であった要因としては、ゲシエンクというゲームに対して構造が複雑、過剰であった可能性が考えられる。加えて、最終的なプレイヤー性能が学習局面数の順位と一致していることから、学習効率が性能差に直結したとも考えられるため、学習局面数を揃えた比較が必要である。

## 4 まとめ

本研究では、ゲシエンクにおけるニューラルネットワーク評価関数の設計を検討した。実験の結果、MLPは勝率25%を超えて勝ち越したが、他のネットワークは勝ち越すに至らなかった。今後の課題として、ネットワーク設計の改善や学習手法の改善などが挙げられる。

## 参考文献

- [1] 楠本 英哲, 野本 瑛二, 松崎 公紀, “カードゲーム「ゲシエンク」における基礎プレイヤーの作成と評価”, 情報処理学会第56回ゲーム情報学研究会, 2025.
- [2] 楠本 英哲, 野本 瑛二, 松崎 公紀, “カードゲーム「ゲシエンク」における深層学習プレイヤーの作成と評価”, 第67回プログラミング・シンポジウム予稿集, 2026.