F-055

# 6Nimmt!におけるカード提出戦略のための期待損失予測
# Expected Loss Prediction for Card Submission Strategies in 6Nimmt!

関 優志† 藤田 悟†
Yushi Seki Satoru Fujita

## 1. はじめに

本研究では、多人数不完全情報ゲーム『6Nimmt!』を対象に、手札の各カードに対する期待損失を予測し、失点リスクの低いカードを提出する AI プレイヤを開発した。場のいずれの行の最後尾のカードよりも小さいカードとそれ以外のカードで異なる特徴量を用い、それぞれに対応したニューラルネットワークを構築した。これにより、状況に応じた準最適なカード提出戦略を実現した。

## 2. 6Nimmt!

6Nimmt!は、2～10 人で遊ぶことのできるドイツ発祥のカードゲームであり、ルールは以下の通りである。

- ゲーム開始時、各プレイヤは 10 枚のカードを手札とする。
- 場に 4 行のカードが配置され、1 行につき最大 5 枚まで置くことができる。
- 全プレイヤが手札からカードを選び、場に伏せて出す。
- カードを表にし、小さいカードから順に、そのカードより小さいカード中で、一番大きなカードの横に並べる。
- 5 枚を超えた行にカードを出した場合、その行の 5 枚を回収して自分のカードを置く。
  - このとき、回収したカードに記載されている牛の数だけ失点する。
- どこにも置くことのできないカードを出すと、自分で指定した 1 行分の牛の数を失点する。
  - この場合、1 行分を回収し、自分のカードを置く。
- 以上の事を 10 ターン行うと 1 ゲーム終了となる。
- 最終的に失点の少ないプレイヤが勝者となる。

以上のルールに従いながら、対戦相手を見つつ、安全に、かつ、失点の少ない行に自分のカードを出していくための戦略を競うゲームである。また、終盤に残すカードを考えながら、序盤戦略を考えることも重要になる。

## 3. 関連研究

Bertschi らの研究[1]では、6Nimmt!を対象に複数の強化学習手法を評価し、AlphaZero をベースとした手法の有効性が示されている。ただし、この研究では、本研究のように、カードの性質に応じて異なる特徴量を用い、手札の各カードについてその期待損失を個別に予測するといったアプローチは採用されていない。

## 4. 提出カードの期待損失予測

### 4.1 概要

6Nimmt!のプレイヤの行動は、次の i.と ii.の２種類に分けられる。

i. 手札から提出するカードを選択する。
ii. 場のどの行にも配置することの出来ないカードを提出したときに、回収する行を選択する。

† 法政大学 情報科学部 Faculty of Computer and Information Sciences, Hosei University.

本研究では、i. の行動について、状態 s のときに行動 a を選択した場合の期待損失を予測する関数 $L(s,a)$ をニューラルネットワークで近似することで、累積失点数を最小化する方策を学習する。また、ii. の行動については、場の 4 つの行の内、最も失点数の少ない行を選択するアルゴリズムを採用した。このとき、失点数が最小の行が複数存在する場合は、行をランダムに選択する。

### 4.2 モデルの構成要素

提案モデルは 2 つのニューラルネットワークで構成される。一つは、場のいずれの行の最後尾のカードよりも小さい手札のカードを提出する場合の期待損失（自ら行を選択して回収することを覚悟した提出戦略の損失）を予測するニューラルネットワークである。もう一つは、場のいずれかの行の最後尾のカードよりも大きい手札のカードを提出する場合の期待損失（行の最後尾への配置を想定した提出戦略の損失）を予測するニューラルネットワークである。両ネットワークは、対象とするカードの性質に応じて異なる入力特徴量を用い、それぞれ独立に学習を行う。

初めに、場のいずれの行の最後尾のカードよりも小さい手札のカードの期待損失を予測するニューラルネットワークについて述べる。このニューラルネットワークは、以下の A から C を入力特徴量として用いる。

A) 各最後尾のカードの中で最小のカードより小さく、対象のカードより大きい未使用のカードの数
B) 現盤面における最小失点数
C) 現在のターン数

本ネットワークは、3 層の全結合層で構成されており、136 次元の入力層、64 ユニットおよび 16 ユニットからなる第 1・第 2 中間層、1 次元の出力層を有する。

次に、場のいずれかの行の最後尾のカードよりも大きい手札のカードについて期待損失を予測するニューラルネットワークについて述べる。このニューラルネットワークは、以下の D から H を入力特徴量として用いる。

D) 対象のカードが配置されるであろう行の空き数
E) 対象のカードが配置されるであろう行の最後尾のカードより大きく、対象のカードより小さい未使用のカードの数
F) 対象のカードより小さい未使用のカードの数
G) 対象のカードが配置されるであろう行以外の行に配置可能なカードの数
H) 現在のターン数

ここで、「配置されるであろう行」は、カード提出前の場において、対象のカードより小さいカードの中で最も大きなカードがある行である。

$$feature_G = \sum_{i \neq j} \min(S_i, R_i) \quad (1)$$

また、特徴量 G)は、場の各行を $i$、対象のカードが配置されるであろう行を $j$、空き数を$S$、最後尾のカードより大きく、配置できる最大のカード以下のカードの内、未使

用のカードの数を$R$とすると、式(1)の通り定義される。

本ネットワークは、3 層の全結合層で構成されており、231 次元の入力層、128 ユニットおよび 32 ユニットからなる第 1・第 2 中間層、1 次元の出力層を有する。

また、両ネットワークにおいて、すべての入力特徴量に対して事前に one-hot エンコーディングを施し、中間層には ReLU 関数を適用した。

### 4.3 学習方法

前節の 2 つのニューラルネットワークは、式(2)で定義される損失関数を用いて学習を行う。この損失関数は、Deep Q-Network [2] における Q 値関数を関数 $L$ に、また即時報酬を失点 $l$ にそれぞれ置き換え、将来の損失は最小のものを選択する形で定義される。また、学習率は 0.001 とする。

$$loss = \left(l_{t+1} + \gamma \min_{a_{t+1}} L(s_{t+1}, a_{t+1}) - L(s_t, a_t)\right)^2 \quad (2)$$

学習の手順は以下の 1 から 4 で構成される。

1. 対戦相手のプレイヤを 5 体用意する。
2. 割引率 $\gamma$ が 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 で設定された提案モデルをそれぞれ用意する。
3. 各割引率の提案モデルに対して以下を繰り返す。
   (i) 手順 2 の提案モデルを順番に 1 つ選ぶ。
   (ii) 対戦相手のプレイヤ 5 体の中から重複を許してランダムに 5 体選ぶ。
   (iii) (i)、(ii)で選んだ計 6 体で 100 ゲーム対戦する。ここで、相手の行動も学習データに含めて学習を行う。
   (iv) (ii)、(iii)を 1,000 回行う。
4. 手順 2 の計 6 体のモデル同士で 71 万ゲーム対戦する。ここで、相手の行動も学習データに含めて学習を行う。

また、本研究では、相手の行動も学習データに含めるため、教師データ作成時において「すべての行の最後尾のカードよりも小さいカード」を提出した場合は、最小失点となる行のうち最も上位に位置する行を回収したとして失点数を概算する。

## 5. 実験

### 5.1 ルールベースプレイヤ

本研究では、比較する対戦相手のプレイヤとして、異なる人が作成した異なるルールベースプレイヤを 10 体用意する。この 10 体の中から、ランダムに 6 体を選び出した対戦を複数回実施し、相対的に弱かったプレイヤを①~⑤とし、強かったプレイヤを⑥~⑩とする。提案モデルの学習には、弱かったプレイヤ①~⑤との対戦結果を用い、評価時には、⑥~⑩の強いプレイヤとの対戦を行って性能を評価する。

### 5.2 評価方法

評価用ルールベースプレイヤ⑥~⑩の 5 体と、各割引率設定の提案モデル 1 体を加えた計 6 体のプレイヤで対戦を行う。ゲーム数は 1 万ゲームとし、各プレイヤの累積失点数を比較する。さらに、1 ゲームごとの失点数に統計的な差が存在するかどうかを検証するため、失点数のデータに対して対応のない 2 群間 t 検定を行う。

### 5.3 結果

実験の結果、累積失点数は表 1 の通りになった。表 2 は、各割引率設定の提案モデルを用いたゲームにおいて、上位 2 プレイヤ間の 1 ゲームあたりの失点数の差を t 検定で検証した結果である。本研究では、帰無仮説として「比較対象のプレイヤ間で、1 ゲームあたりの失点数に統計的な差はない」と設定し、検定における有意水準は 0.05 とする。

**表 1　提案モデル(各割引率 γ)およびルールベース⑥~⑩で対戦したときの累積失点数（単位：点）**

| プレイヤ | 割引率0.3 | 割引率0.4 | 割引率0.5 | 割引率0.6 | 割引率0.7 | 割引率0.8 |
|---|---|---|---|---|---|---|
| ルールベース⑥ | 148253 | 150131 | 152418 | 150508 | 149174 | 150231 |
| ルールベース⑦ | 143393 | 145775 | 145407 | 147478 | 149661 | 147372 |
| ルールベース⑧ | 155186 | 156579 | 155887 | 156007 | 159446 | 158313 |
| ルールベース⑨ | 136507 | 138021 | 138268 | 138619 | 138632 | 139939 |
| ルールベース⑩ | 116711 | 118363 | 118019 | 117367 | 118996 | 118710 |
| 提案モデル(各割引率) | 129803 | 122713 | 120473 | 122438 | **114554** | **116276** |

**表 2　各割引率設定での対戦結果における上位 2 プレイヤ間の 1 ゲームあたりの失点数に対する t 検定結果**

| 提案手法の割引率 | 1位 | 2位 | t値 | p値 |
|---|---|---|---|---|
| 0.3 | ルールベース⑩ | 提案モデル | -9.77 | $1.66\times10^{-22}$ |
| 0.4 | ルールベース⑩ | 提案モデル | -3.31 | $9.35\times10^{-4}$ |
| 0.5 | ルールベース⑩ | 提案モデル | -1.88 | $6.02\times10^{-2}$ |
| 0.6 | ルールベース⑩ | 提案モデル | -3.87 | $1.08\times10^{-4}$ |
| 0.7 | 提案モデル | ルールベース⑩ | -3.39 | $\mathbf{7.06\times10^{-4}}$ |
| 0.8 | 提案モデル | ルールベース⑩ | -1.85 | $6.37\times10^{-2}$ |

表 1 より、提案モデルの割引率が 0.7、0.8 のときに最も少ない累積失点数を記録し、他のすべてのルールベースプレイヤを上回ったことが分かる。また、表 2 より、提案モデルの割引率が 0.8 のときは p 値が$6.37\times10^{-2}$となり、提案モデルと 2 位のプレイヤとの間に有意な差はないことが分かる。一方で、提案モデルの割引率が 0.7 のときは p 値が$7.06\times10^{-4}$となり、提案モデルが 2 位のプレイヤに有意な差で上回ったことが分かる。

結果より、適切な割引率設定において、カードの性質に応じた異なる特徴量を用い、手札の各カードについて期待損失を個別に予測する本手法の有効性が示された。また、割引率が比較的大きい 0.7 と 0.8 で好成績を収めた結果は、6Nimmt! のように将来の損失を正確に予測することが比較的困難な環境においても、将来の損失を重視する戦略が有効であることを示唆している。

今後の課題として、本研究で用いたルールベースプレイヤおよび提案モデルの汎用的な強さを評価する必要がある。また、割引率のみを変化させたモデル同士で対戦させる学習手法の妥当性についてもさらなる検証が求められる。

## 6. おわりに

本研究では、6Nimmt!において、場のいずれの行の最後尾のカードよりも小さいカードとそれ以外のカードで異なる特徴量を用いて、手札の各カードに対する期待損失を予測し、失点リスクの低いカードを提出する AI プレイヤを開発した。実験の結果、適切な割引率設定において本手法がルールベースプレイヤを上回る性能を示した。

### 参考文献

[1] F. Bertschi and J. Mégret, “Understanding Reinforcement Learning with 6nimmt!,” Distributed Computing Group, Computer Engineering and Networks Laboratory, ETH Zürich, (2022).

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing Atari with Deep Reinforcement Learning,” arXiv preprint arXiv:1312.5602, (2013).