

進捗報告

1 GPT4 の出力結果

GPT 4 にカードプールを生成してもらい各役職のデッキを作成してもらった。図 1, 2, 3, 4 に結果を示す。また、

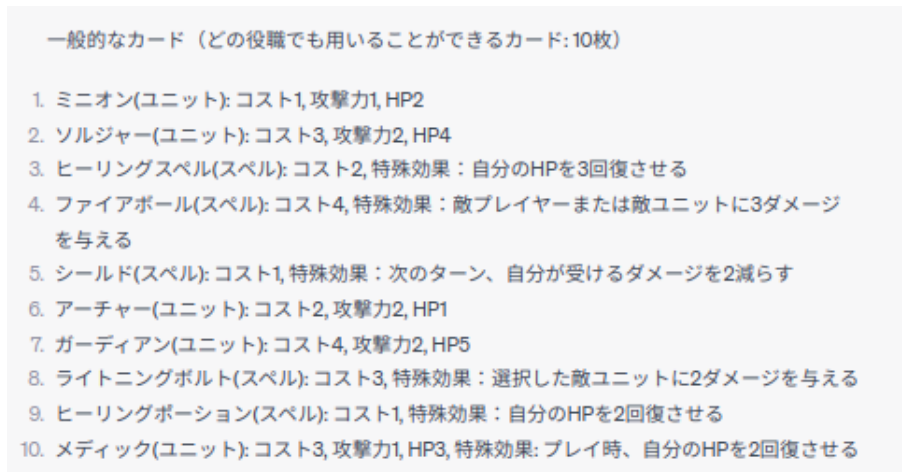


図 1: どの役職でも使えるカード

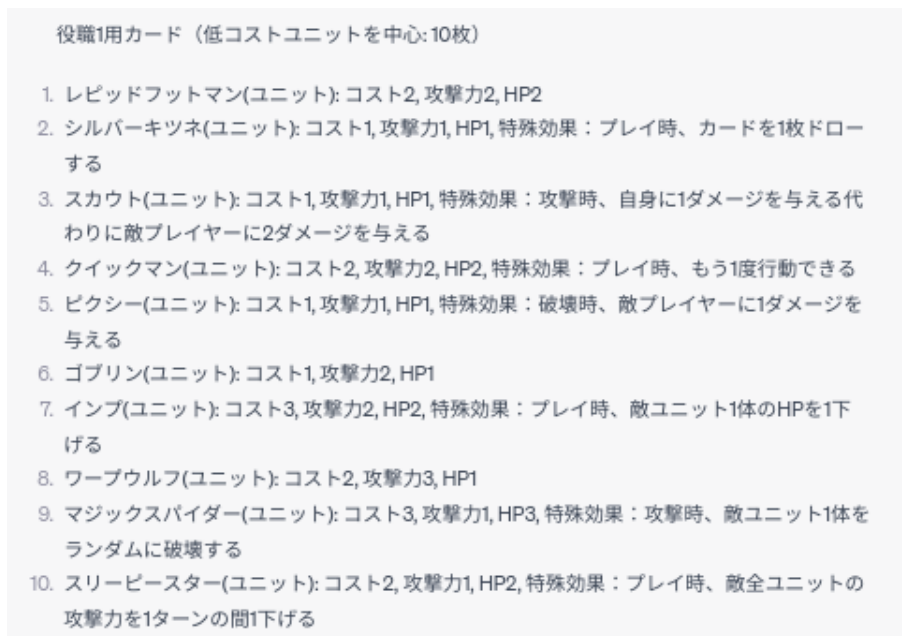


図 2: 低コストユニット中心の役職専用のカード

また、デッキの内訳も決めてもらった。図 5 に出力結果を示す。

役職2用カード（高コストユニットを中心:10枚）

1. ドラゴン(ユニット): コスト7, 攻撃力6, HP6
2. ジャイアント(ユニット): コスト6, 攻撃力5, HP7
3. ゴレム(ユニット): コスト5, 攻撃力4, HP6, 特殊効果: 破壊時、自分のデッキから同名のカードを1枚ドローする
4. デモン(ユニット): コスト6, 攻撃力6, HP4
5. エンタングルドツリー(ユニット): コスト5, 攻撃力3, HP7
6. ティタン(ユニット): コスト7, 攻撃力7, HP5
7. アースエレメンタル(ユニット): コスト6, 攻撃力4, HP8
8. ヘルファイアビースト(ユニット): コスト7, 攻撃力8, HP3
9. スカイガーディアン(ユニット): コスト5, 攻撃力5, HP5
10. クラッキングビースト(ユニット): コスト6, 攻撃力6, HP6, 特殊効果: プレイ時、全ての敵ユニットに1ダメージを与える

図 3: 高コストユニット中心の役職専用のカード

役職3用カード（スペルカードを中心:10枚）

1. メテオ(スペル): コスト5, 特殊効果: 全ての敵ユニットに2ダメージを与える
2. ヒーリングウェーブ(スペル): コスト3, 特殊効果: 自分の全てのユニットのHPを2回復させる
3. サンダーストーム(スペル): コスト4, 特殊効果: ランダムな敵ユニット3体に各1ダメージを与える
4. マジックシールド(スペル): コスト2, 特殊効果: 次のターン、受けるダメージを3軽減する
5. ソウルエクステンジ(スペル): コスト6, 特殊効果: 自分と敵プレイヤーのHPを入れ替える
6. マナフロウ(スペル): コスト1, 特殊効果: 次のターンの初めに、マナを2追加する
7. ブレイズ(スペル): コスト3, 特殊効果: 選択した敵ユニットに3ダメージを与える
8. フォーチュン(スペル): コスト2, 特殊効果: デッキからカードを2枚引く
9. タイムワープ(スペル): コスト5, 特殊効果: 即座にもう一度ターンを取る
10. エネルギーブースト(スペル): コスト4, 特殊効果: 自分のユニット全ての攻撃力とHPを1上げる

図 4: スペルカード中心の役職専用のカード

2 数値実験

図 5 における低コストユニット中心の役職のカードとデッキを実装したため、自己対戦で強化学習のみで戦略を構築する実験をした。

図 6 に学習の様子を示す。1 つのモデルのインスタンスで学習をした。図 ?? に学習中の報酬の推移を示している。

結果として、Test 時の勝率が 0.0 % となり、恐らく何かしらのバグが発生している可能性があると考えられる。急いで他のカードとバグの調査をします。

3 見直すこと

今回のバグの原因として考えられるのがスペルカードの導入です。単体のテストはして問題はないはずなので DQN 側でなにか問題があると考えています。以前まではユニットのみ考えていたため、行動空間や状態空間の定義をし直す必要があると考えています。表 1, 2 に以前までのユニットのみの実験で用いていた状態空間と行動空間の定義を示します。

役職1（低コストユニットを中心）のデッキ

- 一般的なカード: ミニオン x3, ソルジャー x2, ファイアボール x2, ヒーリングスペル x2, アーチャー x3, ライトニングボルト x1
- 役職1用カード: レビッドフットマン x3, シルバーキツネ x3, スカウト x3, クイックマン x3, ゴブリン x2, スリーピースター x2, インプ x1

役職2（高コストユニットを中心）のデッキ

- 一般的なカード: ソルジャー x2, ヒーリングスペル x3, ファイアボール x2, メディック x2, ガーディアン x2, ライトニングボルト x2
- 役職2用カード: ドラゴン x2, ジャイアント x2, ゴレム x2, デーモン x2, エンタングルドツリー x2, クラッキングビースト x1

役職3（スペルカードを中心）のデッキ

- 一般的なカード: ミニオン x2, ファイアボール x3, ヒーリングスペル x3, ライトニングボルト x2, ヒーリングポーション x3, シールド x2
- 役職3用カード: メテオ x2, ヒーリングウェーブ x2, サンダーストーム x2, マジックシールド x2, マナフロウ x3, フォーチュン x2, エネルギーブースト x1

各デッキは自分の役職の特性を活かすように作成しています。しかし、デッキはプレイスタイル、戦略、対戦相手、そして何より楽しさに応じてカスタマイズすることが可能です。このため、これらの例が全ての可能性を網羅しているわけではありません。

図 5: 3 役職のデッキ編成

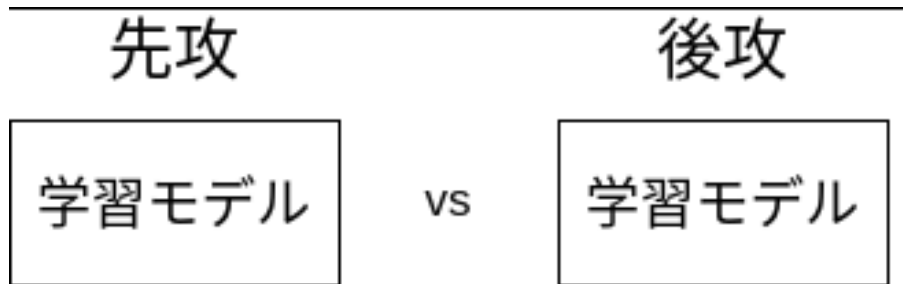


図 6: 学習の様子

また、自己対戦では DQN を用いてきましたが、JSAI の際のぶよぶよの人が用いていた Unity ML-Agent が提供している self-play は PPO と SAC を混合させたような実装をしているようです.[1] まだ実装自体は見つけていませんが、新しいアルゴリズムの実装も考えておきます。

参考文献

- [1] Training in competitive multi-agent environments with self-play. <https://github.com/Unity-Technologies/ml-agents/blob/develop/docs/ML-Agents-Overview.md#training-in-competitive-multi-agent-environments-with-self-play>.

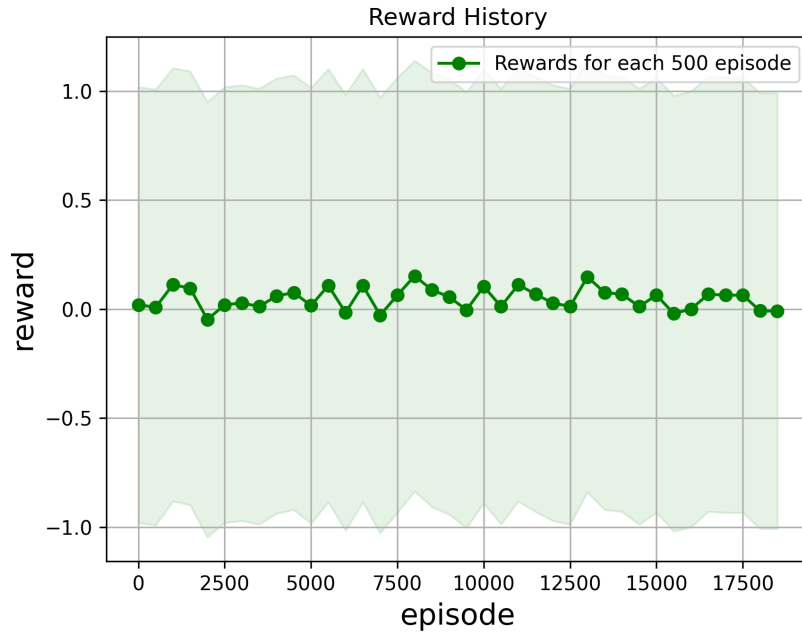


図 7: 学習中の報酬の推移

表 1: 定義した状態空間

| 状態説明 | 次元数 | 最小値 | 最大値 |
|----------------------------------|-----|-----|-----|
| 各プレイヤーの HP | 2 | 0 | 20 |
| 各プレイヤーの マナ | 2 | 0 | 5 |
| 手札 1 ~ 9 の HP, 攻撃力, コスト, 特殊効果 | 36 | 0 | 5 |
| 自盤面 1 ~ 5 の HP と攻撃力 | 10 | 0 | 5 |
| 敵盤面 1 ~ 5 の HP と攻撃力 | 10 | 0 | 5 |
| 自盤面 1 ~ 5 が 攻撃可能かどうか | 5 | 0 | 1 |
| お互いのデッキの 残り枚数 | 2 | 0 | 30 |

表 2: 定義した行動空間

| 行動説明 | 次元数 |
|-----------------------------------|-----|
| 手札 1 ~ 9 を自盤面にプレイ | 9 |
| 自盤面 1 で敵盤面 1 ~ 5 に攻撃 or 敵プレイヤーに攻撃 | 6 |
| 自盤面 2 で敵盤面 1 ~ 5 に攻撃 or 敵プレイヤーに攻撃 | 6 |
| 自盤面 3 で敵盤面 1 ~ 5 に攻撃 or 敵プレイヤーに攻撃 | 6 |
| 自盤面 4 で敵盤面 1 ~ 5 に攻撃 or 敵プレイヤーに攻撃 | 6 |
| 自盤面 5 で敵盤面 1 ~ 5 に攻撃 or 敵プレイヤーに攻撃 | 6 |
| ターンエンド | 1 |

```
win_sum
0
loss_sum
10000
win rate
0.0
episode_number
18720
学習おわり
```

図 8: 検証時の勝率 (恐らくバグ)