未知の場を短時間で学習するドミニオンAI

田中開士[1,a)]　橋本剛[1]

概要：近年，完全情報ゲームや一部の不完全情報ゲームでは人間を超えたが，多くの複雑な不完全情報ゲームでは強い AI は存在しておらず研究が盛んである．複雑な不完全情報ゲームの一つであるドミニオンはターン内での逐次選択が多く深層強化学習に向いている．その一方，ゲームごとに別種のカードの組み合わせを用いるため全ての場を事前に学習することは難しい．そこで，未知の場を学習する際は既存の学習した場の中から類似性の高い場の学習モデルを用いることで学習時間の短縮が可能なのではないかと考えた．既存の学習モデルをスタートモデルとして利用するため，変更するカードは効果のみを変更することにより学習モデルにあたかもこれまでと同じ場を学習しているかのように錯覚させる Learning by Deceive Method（LDM）を提案した．LDM を用いた実験により大幅な学習時間短縮が確認できた．

Dominion AI to learn unknown places in a short time

Tanaka Kaito[1,a)]　Hashimoto Tsuyoshi[1]

Abstract: In recent years, AI has surpassed humans in perfect information games and some imperfect information games, but there is no strong AI for many complex imperfect information games, and research is active in this area. Dominion, one of the complex incomplete information games, is well suited for deep reinforcement learning due to the large number of sequential selections within a turn. On the other hand, it is difficult to learn all fields in advance because each game uses a different combination of cards. Therefore, we considered that it would be possible to shorten the learning time by using the learning model of a field with high similarity among the existing learned fields when learning an unknown field. We proposed the Learning by Deceive Method (LDM), which uses an existing learning model as a starting model and changes only the effect of the cards to make the learning model think that it is learning the same place as before. Experiments using LDM showed a significant reduction in learning time.

1. はじめに

完全情報ゲームでは囲碁で AlphaGo[1] が人間を超えるなど大きな成果が上がっている．一方，不完全情報ゲームでは麻雀で Suphx[2] が人間のトッププレイヤーに匹敵する成績を達成したとされているが多くのゲームで強い AI は存在せず，近年研究が盛んである．本研究では複雑で膨大な盤面データを持ち，カードを用いる不完全情報ゲームのドミニオンを題材とする．ドミニオンは世界三大ボードゲームの一つで，拡張セットを含めて１４種類が販売され，オンライン対戦サイトを用いた日本大会も開かれる人気ゲームである．ドミニオンはゲームで使用するカードの効果やルールにより内容が複雑であり人間に勝てる AI を作成することは難しい．オンラインサイトにも対戦用 AI は存在するが人間を超えるには至っていない．ドミニオン AI の先行研究に Yang らによる固定のカードでの深層強化学習による AI が開発されている．ドミニオンはターン内における逐次選択が多く，また先を見通した戦術が必要であり，これらは強化学習に適していると考えられる．本研究では深層強化学習を用いた強いドミニオン AI の作成を目標とする．しかし，ゲームごとに別種のカードの組み合わせを用いるドミニオンにおいて全てのパターンを事前に学習することは難しい．そこで実用的な対戦用 AI として，ゲーム開始時にセットした場に対して短い時間で学習できる AI の実現を目指す．強い人間プレイヤーは未知の

[1] 松江工業高等専門学校
National Institute of Technology, Matsue College
[a)] s2212@matsue-ct.ac.jp

場でも過去の似たような場の経験をもとに素早く戦術をたてる．AI も過去の経験として既存の学習モデルを利用し，短い時間で学習する事が可能なのではないかと考えた．そこで本研究では事前にいくつかの固定の場を学習し，そのモデルを利用することで短い時間で未知の場の学習を行うシステムを検討し，実験により強さを評価する．

2 章ではドミニオンについて述べる．3 章では深層強化学習について述べる．4 章では短時間での学習システムについて述べる．5 章では固定の場での実験について述べる．6 章では未知の場での実験について述べる．7 章では未知の場での実験結果からの考察と問題解決案について述べる．8 章では名義のみを変更した未知の場での章実験について述べる．9 章では名義のみを変更した未知の場での実験結果からの考察を行う．10 章ではまとめを行う．

2. ドミニオンについて

2.1 ドミニオンとは

本研究で用いるドミニオンは，アメリカ発祥の自分の領土（デッキ）をゲーム中の行動で強化していくことが目的のデッキ構築型のボードゲームである．作者はドナルド・X・ヴァッカリーノであり 2008 年秋にアメリカのリオグランデゲーム社より発売された．アラカルト・カードゲーム賞，ドイツゲーム賞 2009，ドイツ年間ゲーム大賞の世界的なゲーム賞にて史上初の三冠を成し遂げたゲームである．日本国内でも人気であり，ホビージャパンより日本語版が発売されており，ドミニオンオンラインというオンラインサイトを用いた日本大会が開催されている．

2.2 ドミニオンのルール

ドミニオン開始の準備として，場 (図 1) に銅貨（1 金）・銀貨（2 金）・金貨（3 金）の「財宝」カード，屋敷（1 点）・公領（3 点）・属州（6 点）の「勝利点」カード，-1 点の「呪い」カード，10 種類の王国カードの山を置く．王国カードは，基本カード（銅貨，屋敷，属州，呪い，など）以外のカードほぼすべてのことをいう．この 10 種類の王国カードはゲームごとにセットから公式サプライから選んだりランダムに選んだりして決める．この王国カードの選択により多様なゲーム展開を見せるのがドミニオンの特徴である．

各プレイヤーは銅貨を 7 枚，屋敷を 3 枚受け取り，山札として使用する．山札から，プレイヤーが 5 枚の手札を引いてゲームを開始する．銅貨 7 枚を元手に任意のカードを購入・獲得することで，デッキを構築していくことになる．購入・獲得したカードは捨て札に置かれ，すぐに利用することはできない．ターンの終了時，使ったカードや余った手札はすべて捨て場に捨てて，新たに山札から 5 枚引いて手札とする．山札がなくなると，捨て札をすべて切り混ぜて新たな山札とする．この繰り返しにより得たカードは山札に取り込まれ，デッキが構築される．各プレイヤーの手番はおもに下記の 3 つのフェイズで構成される．各フェイズを終了したら，他のプレイヤーに手番を渡す．各フェイズは独立していて，購入フェイズの後にアクションフェイズに戻るようなことはできない．

・Action フェイズ
手札にあるアクションカードを使用できる．アクションカードにはそれぞれ能力がある．アクションカードを場に出すことで，書かれている効果を上から順に処理していく．アクションを使えるのは 1 ターンに 1 回までだが，カードの効果によって，追加のアクションを行うこともできる．手札の枚数を増やしたり，お金を生み出したりするアクションカードもある．

・Buy フェイズ
手札の財宝カードを場に出すことで，お金が生み出される．各カードにはそれぞれ値段（コスト）が設定されており，コスト分のお金を消費することで，場にあるカードを購入できる．購入により獲得したカードは捨て札に置く．購入に使用されなかった分の，アクションや財宝により発生したお金はターン終了時に消える．購入は 1 ターンに 1 回までだが，カードの効果によって，追加の購入を行うこともできる．

・Clean up フェイズ
このターン使用したことで場に出ているカード，及び手札のカードを全て捨て札に置く．その後，山札から 5 枚のカードを引き手札とする．山札がなくなり引けなくなった場合は捨て札をシャッフルし，山札にして残りを引く．

最終的には，場の属州のカードの山か，その他のカードの山の 3 つが無くなった時点でゲームを終了する．このとき各プレイヤーのデッキに含まれる勝利点の合計を計算し，最も勝利点の高いプレイヤーの勝利となる．拡張セットの植民地も属州と同様に無くなった時点でゲームを終了する．

ドミニオン基本セット内のカード一覧

・銅貨 ・銀貨 ・金貨 ・屋敷 ・公領 ・属州 ・呪い
・礼拝堂 ・堀 ・家臣 ・工房 ・商人 ・前駆者
・地下貯蔵庫 ・鍛冶屋 ・村 ・改築 ・祝祭 ・民兵
・密猟者 ・玉座の間 ・議事堂 ・役人 ・庭園 ・市場
・衛兵 ・研究所 ・鉱山 ・金貸し ・書庫 ・山賊
・魔女 ・職人

より詳しいルールは wiki[3] を参考とする．

3. ドミニオンと深層強化学習

3.1 深層強化学習

深層強化学習は Deep learning を用いた強化学習であり，近年多くのゲーム AI 開発に用いられている．完全情報ゲームである囲碁では Deep Mind 社が開発した AI「AlphaGo」が当時の世界チャンピオンに勝利しており AI が人間を超え

図 1　ドミニオン開始時の場

たとされている [1]．また不完全情報ゲームの麻雀では 2019 年にマイクロソフト社によって開発された AI「Suphx」が人気麻雀ゲーム「天鳳」にて最高段位の 10 段を達成している [2]．これら二つのの AI には部分的に深層強化学習が用いられている．Yang ら [4] はカードが固定された二人用のゲームにおいて深層強化学習を利用しドミニオンのゲーム AI を解発した．本研究では固定されていないカードの場において AI の開発を行う．

3.2　未知の場の強化学習への適応

ドミニオンはカードの種類の豊富さや複雑なゲームルールから逐次的な意思決定と先を見通した行動が求められるゲームであり，これらの選択にはプレイヤーの経験が十分に必要なゲームと考えられる．これらの選択をスムーズに行うのに深層強化学習は適していると考えた．しかし，ドミニオンには 2.2 章で説明したとおり下記のようなゲームを複雑にするルールがある．

(1) デッキ構築型ゲームで，ゲーム内の行動選択によりプレイヤーが使用するカードの数や種類が変化
(2) ゲームごとに使用する王国カード 10 種類を任意の方法で選んでゲームを開始する
(3) カードの種類は基本セットでは 26 種類，拡張すべてを含めると 500 種類以上
(4) カードごとに効果が異なる特に強化学習への適応を難しくしているのが (2) と (4) のルールである．強化学習では決まった盤面において反復的に学習することにより AI モデルのアップデートを行うが，(2) のルールによりゲームごとに別種のカードの組み合わせを用いる場合もあるため学習時に未知のカードにより未知の場が形成される．これら未知の場では強化学習を行うことが難しく，また決まった戦術を取ることも難しいと考えられる．また (4) のカード効果の違いも決まった戦術を取ることを難しくしており，1 枚のカードが戦術に大きく影響を及ぼすようなものも存在する．これらの要素から事前に決まった戦術をとるのでは十分な勝率を上げるのは難しく，またゲームごとの盤面の変化が強化学習への適応難易度をあげている．

4.　学習戦術選定による高速化

ドミニオンは 3.2 章のような要因からそのままのルールのもと強化学習への適応をすることは難しいと考えられる．しかし，人間はドミニオンをプレイする際も未知の場に対してある程度の適応がなされている．そこで人間が未知の場に対して思案する時と同じ形となるようにシステムの設計を行う．通常人間はドミニオンをプレイする際カードのセットが行われた後このゲームでプレイヤーがとる戦術を考える．その際，過去に体験したことのある似たような場やカードから戦術だてを行う．似たような場には似た戦術が有効である場合が多いためであり、このような経験から考えられた戦術をゲーム中に修正しながら人間は未知の場に適応しているのだと考えられる．このような考え方を AI 学習システムに合わせて考えると，事前にいくつかの固定の場において学習を行った戦術モデルを用意しておき，未知の場が与えられた際に用意された戦術モデルで学習した場と未知の場との類似性について調べる（図 2）．類似性の高かった戦術モデルをスタートモデルとして学習を開始する（図 3）ことで，未知の場に対して一から学習することで生じる膨大な学習時間を削減しつつ，未知の場に適応した学習モデルの生成が可能なのではないかと考えられる．

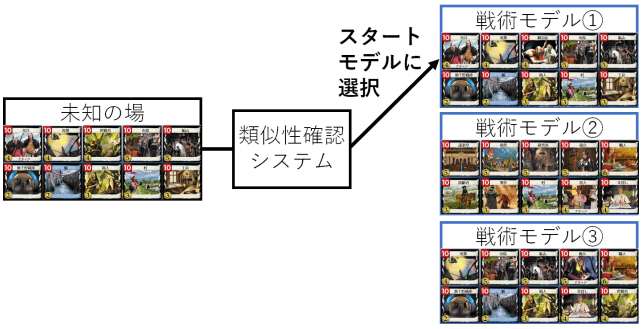


図 2　類似性確認システムによる未知の場と類似性の高い場の選択

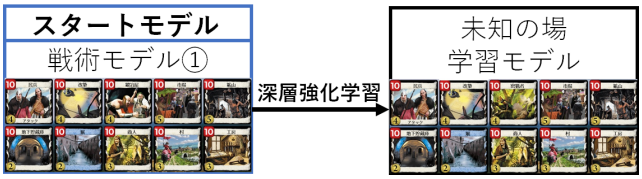


図 3　スタートモデルによる未知の場の学習

5.　固定の場での実験

5.1　実験概要

予備実験として固定されたカードの場の学習を行い深層強化学習により戦術モデルの生成が可能であるか検証する．Lowman のドミニオンプログラム [5] と Keras[6] を用

いて深層強化学習の環境を構築した．Deep learning の構成を下記の図 4 のように設定した．各ターンの行動選択と盤面データを記憶し，ゲーム終了時，勝敗と合わせ記録する．学習時は，これらのデータを入力とする．出力は行動選択とする．報酬は勝利 1，敗北-1，引き分け 0 が与えられる．学習時は 1/3 の確率でランダムに行動し，2/3 の確率で学習モデルの行動をとる．10 game を 1 iteration とし iteration 終了時に学習モデルをアップデートする．その後ヒューリスティックな AI である BigMoneyAI (以下 BM) と対戦し，対戦結果を記録した．BM は下記の行動をとる．

・アクションカードは使用しない

・カード購入時に持ちコインが 3 未満なら購入しない，6 未満なら銀貨，8 未満なら金貨，8 以上なら属州を購入する．

実験では固定の場として下記に記した公式サプライで設定されている「はじめてのドミニオン」を用いる．BM は著者と固定の場で対戦した結果 21 ターン，勝利点が 36-21 で著者が勝利した．

「はじめてのドミニオン」の場

・銅貨 ・銀貨 ・金貨 ・屋敷 ・公領 ・属州 ・呪い
・堀 ・工房 ・商人 ・地下貯蔵庫 ・鍛冶屋 ・村
・改築 ・民兵 ・市場 ・鉱山

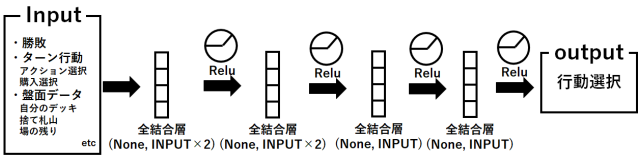


図 4 Deep Learning の構造

5.2 結果と考察

BM との対戦による勝率を図 5 に示す．iteration の増加に伴い勝率の上昇が見られた．特に 10698 iteration と 10725 iteration で BM に対して勝率が 80 %に到達している．表 1 にそれぞれの BM との対戦結果を示す．これらの結果から深層強化学習によってドミニオンの学習が可能であることが分かった．しかし，学習時間としては 1 iteration に約 2 分 30 秒，10725 iteration で約 400 時間（26812.5 分）以上かかっており，到底実用的な対戦用 AI とは言えない状態である．また図 6 にそれぞれの BM との対戦時の各種平均獲得カードを示す．このカード群から属州という勝利点効率の高いカードを得ることができていると分かる．

表 1 固定の場の BM との対戦結果

iteration	Win	Draw	平均勝利点	学習時間
10697	80 %	0 %	33.6 点	26742.5 分
10725	80 %	20 %	32.1 点	26812.5 分

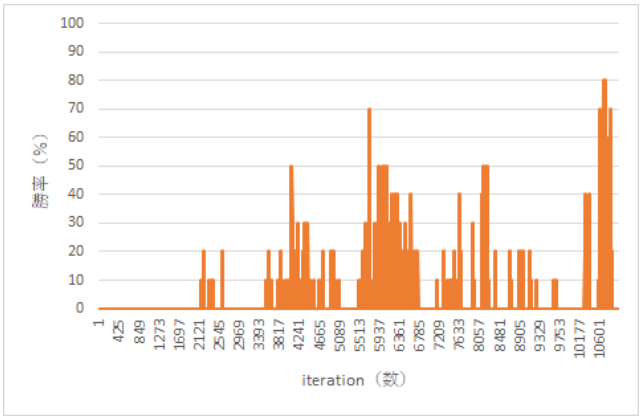
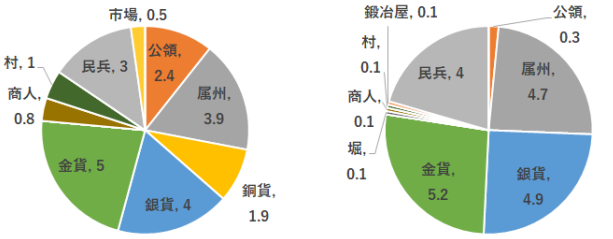


図 5 固定の場での BM との iteration ごとの勝率



(a)固定の場，10697 iteration　　(b)固定の場，10725 iteration

図 6 固定の場での BM との対戦時の各種平均獲得カード数

6. 未知の場での実験

未知の場において類似性の高い場を学習したモデルをスタートモデルとして短時間で学習することは可能であるのか実験を行う．

6.1 実験 1

本実験では類似性の高い場とは同盟カードの多い場と解釈して実験を行う．そこで固定の場の学習においてスタートモデルとなる学習モデルができている「はじめてのドミニオン」の場から「鍛冶屋」を「密猟者」に変更した場（図 7）を用いて実験を行う．固定の場の実験により BM に対して勝率 80 %を超えた 10725 iteration 時のモデルをスタートモデルとし固定の場と同様の学習と記録を行う．

図 7 実験 1 で用いる未知の場

6.2 実験 1 の結果

結果，勝率は 0 %のままであった．また図 8 に示した通り，平均勝利点も二桁に到達しないままであった．

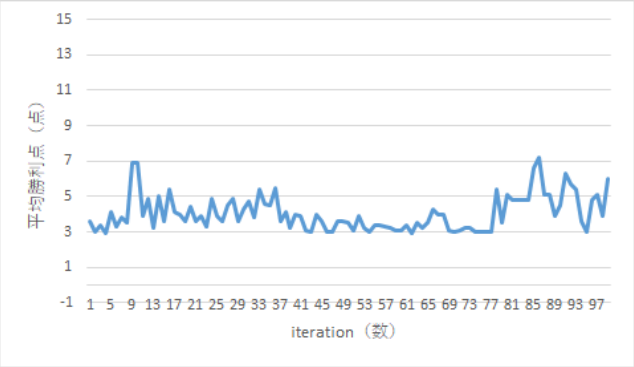


図 8　実験 1 での BM との iteration ごとの勝利点

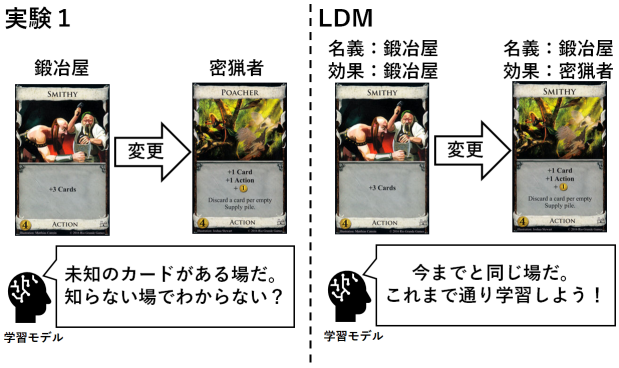


図 10　Learning by Deceive Method（LDM）

7.　LDM

7.1　実験 1 の考察

事前の想定では実験 1 において鍛冶屋を密猟者に変更してもスタートモデルは鍛冶屋の獲得頻度は低いため戦術に影響は少なく，iteration 数が少なくてもでも BM に対して勝利することが予想されていた．しかし結果は思ったように学習がなされたいないことが分かった．そこで BM との対戦時の獲得カードを確認すると固定の場で獲得していたカード（図 6）とは違うカード群を獲得している（図 9）ことが分かった．これはスタートモデルに密猟者という未知のカードが入ったことによりこれまでの学習が生かされていないためこのような結果になったのではないかと考えられる．

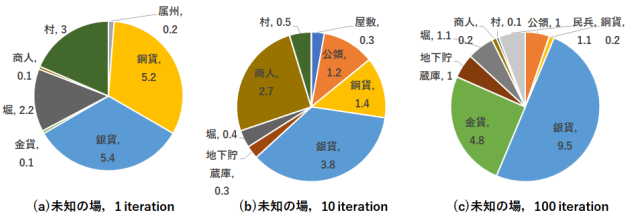


図 9　未知の場での BM との対戦時の各種平均獲得カード

7.2　LDM

未知のカードが存在していると既存の学習モデルをスタートモデルとして利用できていないようである．この問題の解決策として名義は同じカードとして扱いプログラムの内部的な効果のみを変更することにより，学習モデルに入力される部分を同じにすることであたかもこれまでと同じ場を学習しているかのように錯覚させる（図 10）ことで未知の場の学習が可能なのではないかと考える．そこで，このような学習モデルをだまして未知の場を学習させる方法を提案し，Learning by Deceive Method（LDM）と名づける．

8.　LDM による未知の場での実験

8.1　実験 2

実験 1 と同様の実験を行う．その際 LDM に則りプログラムコード内の鍛冶屋の効果を密猟者の効果に変更し，名義上は鍛冶屋であるが密猟者の効果のカードを使用した場合の学習と記録を行う．

8.2　実験 2 の結果

勝率を図 11 に，勝利点を図 12 に示す．結果，勝率は上がった．1 iteration で勝率 50 %を記録すると，2 iteration で 60 %，19 iteration で 70 %となった．勝利点も高い結果を記録している．また各 iteration での各種平均獲得カードについては図 13 に示す．

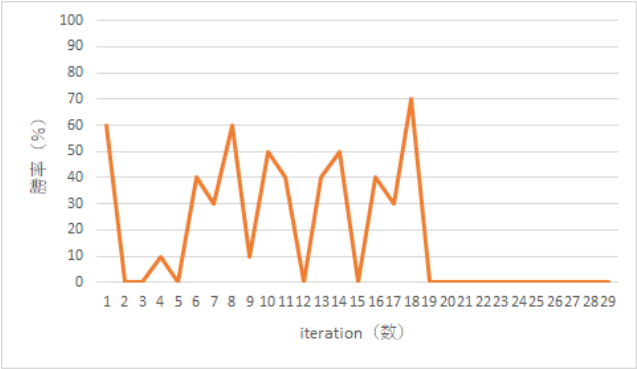


図 11　実験 2 での BM との iteration ごとの勝率

8.3　実験 3

実験 1 と実験 2 ではスタートモデルにおいて獲得頻度の低い鍛冶屋を変更して実験を行った．実験 3 ではスタートモデルにおいて獲得頻度の高い「民兵」を「役人」に変更した場（14）を用いて実験を行う．実験 2 と同様 LDM に則り，民兵の効果を役人に変更して学習と記録を行う．

8.4　実験 3 の結果

勝率を図 15 に，勝利点を図 16 に示す．結果 21 iteration

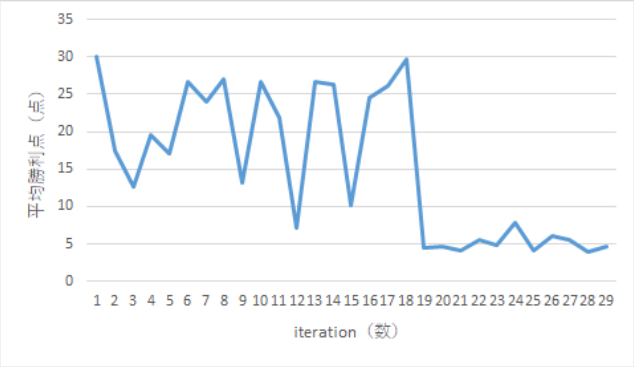


図 12 実験 2 での BM との iteration ごとの勝利点

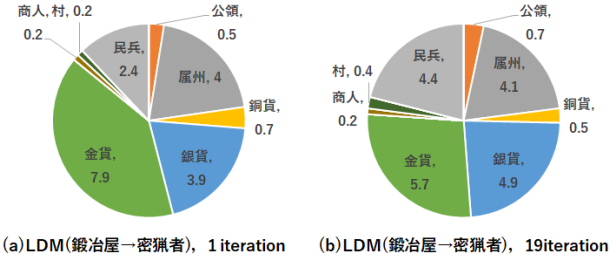


図 13 LDM による未知の場での BM との対戦時の各種平均獲得カード

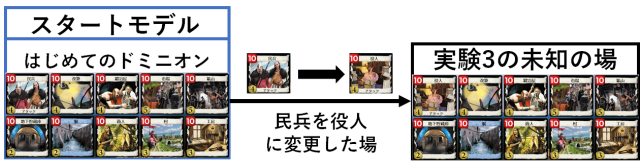


図 14 実験 3 で用いる未知の場

で勝率 10 %を記録したが，その後は長い間勝率の向上が見られなかった．その後 242 iteration で勝率 20 %，245 iteration で勝率 30 %となったがそれ以上の改善は見られなかった．

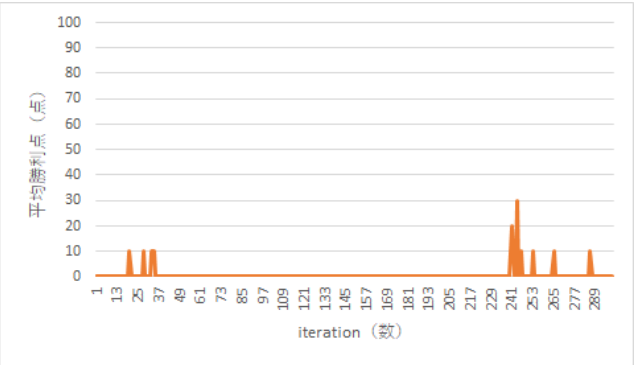


図 15 実験 3 での BM との iteration ごとの勝率

9. 考察

実験 1 を踏まえて実験 2 を行ったところ想定していたように密猟者の効果のカードを鍛冶屋と錯覚した状態で変更

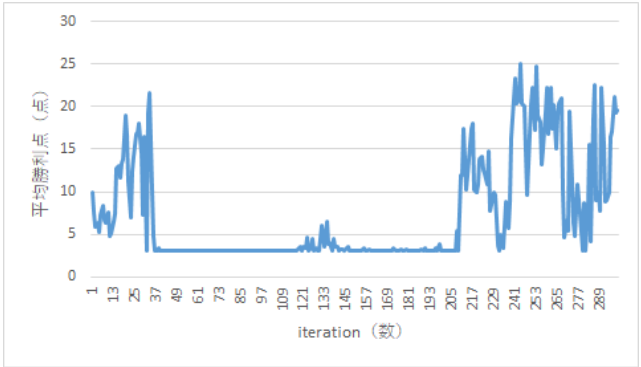


図 16 実験 3 での BM との iteration ごとの勝利点

前の場と同じ場だと認識しての学習がなされた．実験 2 の BM との対戦時の獲得カード（図 9）を確認してみても固定の場で獲得していたカードと同様のカード群を獲得していたことがわかった．これにより LDM を用いることにより固定の場で学習したモデルをスタートモデルとして未知の場を学習することができることが分かった．また少ない iteration で高い勝率を収めたことから固定の場の実験でみられたような一から学習することによる勝率の上がらないままの膨大な学習時間を短縮することができていると考えられる．しかし，19 iteration であっても約 50 分弱の時間を有しており，実用的な対戦 AI とするには難しいと考えられる．実験 3 ではスタートモデルにおいて獲得頻度の高いカードを変更した場合の実験を行った．結果，実験 2 のような少ない iteration で高い勝率を得ることはできなかった．けれども固定の場の実験結果と比べると十分に少ない iteration で BM に勝利することができるようになっているといえる．このことからスタートモデルにおいて獲得頻度の高い，すなわち重要度の高いカードが変更された場合は類似性が高い場が最適なスタートモデルであるかについては引き続き実験が必要であることが分かった．これらの結果から今回は類似性の高さを同じカードを多く使用している場と解釈して実験を行ったが，スタートモデルにおいて重要度の高いカードが変更させた際には別の指標を用いてスタートモデルの選択を行う必要性が考えられる．

10. おわりに

本研究にて未知の場に対して類似性の高い場を学習したモデルをスタートモデルとすることにより，未知の場を一から学習するよりも時間を大幅に短縮することが出来ることが分かった．またその際には LDM を用いる事により未知の場であっても別の場の学習モデルを流用して学習を行うことが可能であることがわかった．しかし，実用的な対戦用 AI とするには学習時間が現状でも十分必要であり，スタートモデルにおいて獲得頻度の高いカードが変更された場合は高い勝率を達成できない場合がある．また学習済みの場が一定量必要であり，それらを用意するにも十

分な時間を要するため自己対戦や学習の高速化が課題である．また 5 や 11 の後半部分において勝率が 0 %の間が多く存在することについて学習の自己対戦時のランダム行動率が 1/3 と大きい状態で続けていたことが問題であり，学習方法についても改善すべき部分が存在している．今後の研究として本研究では行わなかった「はじめてのドミニオン」の別のカードを変更した場，学習済みである「改善」や「デッキトップ」の場のカードを変更した場においても同様の実験を行い効果の検証を行うとともに，自己学習の速度向上，学習方法の見直しを行っていく．

参考文献

[1] Silver, D., et: Mastering the game of go with deep neural networks and tree search, Nature, Vol. 529, pp. 445-446 (2016)
[2] Junjie Li, Sotetsu Koyamada, Qiwei Ye, Guoqing Liu, Chao Wang, Ruihan Yang, Li Zhao, Tao Qin, Tie-Yan Liu, Hsiao-Wuen Hon: Suphx: Mastering Mahjong with Deep Reinforcement Learning, Arxiv Mar 2020. [Online]. Available: https://arxiv.org/abs/2003.13590 (2020)
[3] Dominion Strategy Wiki, https://wiki.dominionstrategy.com/ （アクセス日 2023.10)
[4] Hung-I Yang , Yu-Chi Kuo: An AI for Dominion using Deep Reinforcement Learning (2019)
[5] Corey Lowman:python implementation of the Dominion board game – GitHub , https://github.com/coreylowman/dominion （アクセス日 2022.1)
[6] Keras , https://keras.io/ （アクセス日 2023.10)