Vol.2013-MPS-96 No.4
Vol.2013-BIO-36 No.4
2013/12/11

コンピュータ大貧民における高速な相手モデル作成と精度向上

伊藤祥平[†1]　　但馬康宏[†1]　菊井玄一郎[†1]

UEC コンピュータ大貧民大会ではモンテカルロ法を用いたクライアントが優勝している. そこでプレイアウト中の相手着手を実際の着手に近づけることでモンテカルロ法によるクライアントの強化を考える．本研究ではゲーム中の実際の相手着手を学習する方法としてナイーブベイズを用いる．これにより高速な相手のモデル化を行う．さらに、学習素性の工夫により精度の向上を行った．この結果，作成されたモデルの精度は過去の優勝クライアント snowl に対し，4 割程度の近似ができた.

The Accuracy Improvement with The fast Opponent Modeling in The Computer DAIHINMIN

SHOUHEI ITO[†1]　YASUHIRO TAJIMA[†1]
GENICHIRO KIKUI[†1]

Monte-Carlo method is also useful for DAIHNMIN and the client using this method has won the UEC computer DAIHINMIN tournament. We try to accelerate the strength of Monte-Carlo method by making effective opponent models which are close to the real opponents' moves. Stronger opponent models, more effective playouts our client has. We use Naive Bayes as the learning method to modeling the opponents. This method is one of the fastest algorithm for learning and classification. In addition, its accuracy is enough to modeling the opponents. In this paper, we show two modeling by Naïve Bayes. The first method is the simple modeling , and the second is improved the move data structure. The accuracy is approximately 40% by our improved method to model snowl which is the champion client in 2010.

1. はじめに

UEC コンピュータ大貧民大会では近年，モンテカルロ法を用いたクライアントが優勝している．モンテカルロ法での行動決定では，採り得る各行動に対して試合終了までのシミュレーション(プレイアウト)を複数回行うことで行動を決定する．そこでプレイアウト中の相手着手を実際の着手に近づけることでモンテカルロ法によるクライアントの強化を考える[1]. そのためにはゲーム中の相手着手をゲーム中に学習し，相手の着手と似たような着手を行うモデルを作成する必要がある.

作成するモデルは 1 回の着手決定で数千回行われるプレイアウト中で使用するため, 高速に判別できる必要がある. また，ゲーム中に相手の着手を学習することになるため，高速であり，少ない学習データでもある程度の精度が必要となる.

本研究ではゲーム中の実際の相手着手を学習する方法としてナイーブベイズを用いる．これにより高速な相手のモデル化を行う．さらに，学習素性の工夫により精度の向上を行った．この結果，作成されたモデルの精度は過去の優勝クライアント snowl[2]に対し，4 割程度の近似ができた.

2. UEC コンピュータ大貧民

大貧民では地方によってルールに若干の違いがある．本研究では UEC コンピュータ大貧民大会のルールに従うものとする. 以下に UEC コンピュータ大貧民大会のルールを簡単に説明する．ただし，基本的な部分の説明は省く.

- 使用カード : ジョーカー1 枚を入れた計 53 枚.
- パス : いつでも可能．ただしパスした場合，場が流れるまで順番は回ってこない.
- 1 ラウンドの開始 : ダイヤの 3 を持っている人が 1 番最初にカードを出す権利を持つ．必ずしもダイヤの 3 を出す必要はない.
- 上がり方 : どんなカードでも上がることができる.
- 複数枚同時出し(ペア) : 同じ数字カードを複数枚同時に場に出すことができる.
- 階段(連番) : 同じマークでつながった数字が 3 枚以上であれば，同時に場に出すことができる.
- 革命 : 4 枚以上の複数だし，または 5 枚以上の階段を場に出すと発生する.
- ジョーカー : ジョーカーは単独で出せ革命関係なく最強のカードである. 他のカードと組み合わせた場合はその組み合わせに必要なカードとして出すことができる.
- スペードの 3 : 基本的には単なる 3 のカードである. ただし，ジョーカーが単独で出ている場合，ジョーカーよりも強いカードとして出すことができる. この場合，出した後は場が流れる.
- 8 のカード : 出すと場が流れる.
- しばり : 同じマークの組合せが 2 回出ると発生する.

†1 岡山県立大学
Okayama Prefectural University.

Vol.2013-MPS-96 No.4
Vol.2013-BIO-36 No.4
2013/12/11

しばりが発生した場合，以降場が流れるまで同じマークの組合せのみ出すことができる．このしばりはジョーカーを使って発生させることも可能である．

3. 関連研究

モンテカルロ法を用いたクライアントには 2009 年の UEC コンピュータ大貧民大会で優勝した fumiya[3]，2010 年に優勝した snowl，2011 年に優勝した crow[4]がある．

fumiya ではプレイアウトの結果として，プレイアウト終了時の自分のランクを使用する．大貧民であれば 1 であり，大富豪であれば 5 となる．プレイアウト時の着手はランダムで行う．ただし，場にカードがないときの着手決定ではパスを除く．また，プレイアウトを行う可能手の決定には UCB1-TUNED を用いており，プレイアウト時の相手手札はランダムで決定している．

snowl は fumiya を改良したクライアントである．プレイアウトには fumiya の着手を学習データとした Policy Gradient Simulation Balancing を用いて決定し，プレイアウト時の相手手札は BRATTERY-TERRY を用いて推定している．また，モンテカルロによる探索を行う前に必勝手探索を行っており，そこで必勝手が見つかればモンテカルロによる着手決定は行わない．

crow は snowl を改良したクライアントである．プレイアウトの結果をプレイアウト終了時のプレイヤーのランクの推移にも着目し，プレイアウトの結果を差分学習により改善している．

また，クライアントの着手分析の研究には次のような研究がある．文献[5]ではクライアントの着手に対してクラスタ分析により大貧民クライアントの客観的な分類化を実現している．文献[6]ではそれぞれのクライアントの着手の一致率を調べている．この文献では default と snowl の一致率はゲーム序盤では 4 割程度，中盤では 5 割程度となっている．ここで default とは UEC コンピュータ大貧民大会で配布されているクライアントである．このクライアントは一番弱い可能手を優先して着手する．また，手札の組合せを崩さない着手を行う．手札の組合せを崩すとは手札内に着手のカードを使って着手より多く出す組み合わせがある時に，着手をしてその組み合わせがなくなることを表す．

4. 提案手法

プレイアウト中の相手着手を実際の着手に近づけることでモンテカルロ法によるクライアントの強化を考える．そのためにはプレイアウト中に使用する相手モデルは実際の対戦相手と似たような着手を選択する相手モデルが必要がある．そこで本研究ではゲーム中に相手が実際に行った着手をゲーム中に学習する．よって相手モデルは学習と推定に時間が掛からず，尚且つ少ないデータでも推定することができる必要がある．本研究ではナイーブベイズによる相手モデルを作成した．相手モデルは相手 1 人分の着手のみ学習するものとする．

ナイーブベイズによるプレイアウト中での相手モデルによる着手決定では，その時の全可能手Aについて手札$\boldsymbol{d}$の時，着手A $\in l$を出す確率$P(l|\boldsymbol{d})$が最大のものを選択する．ここで，着手lは着手の強さを表すものとする．$P(l|\boldsymbol{d})$を以下で示すベイズの定理により求める．

$$P(l|\boldsymbol{d}) = \frac{P(l|\boldsymbol{d})P(l)}{P(\boldsymbol{d})}$$

ここで，$P(\boldsymbol{d})$はその時点での全ての着手lに対し，同じなので分子のみ考える．n枚の手札$\boldsymbol{d}$は手札のカードの強さd_iの集合

$$\boldsymbol{d} = \{d_1, d_2, \cdots, d_i, \cdots, d_n\}$$

とし，$P(l|\boldsymbol{d})$を以下の式で表す．

$$P(l|\boldsymbol{d}) \propto P(l)P(d_1|l)P(d_2|l)\cdots P(d_n|l)$$

また，大貧民は場に出ているカードによって可能手が極端に制限される．そこで階段用と複数・単独用のそれぞれで相手モデルを作り，場の状況によって相手モデルを使い分けるようにする．

カードの強さは表 1 のようにする．また着手の強さは着手に使われているカードの中で最弱カードの強さとする．

表 1. カードの強さ

カード	3	4	～	A	2	ジョーカー
通常時	1	2	～	12	13	14
革命時	13	12	～	2	1	14

さらに本研究ではこのナイーブベイズの素性を改良した改良版のナイーブベイズによる相手のモデル化を行った．改良版のナイーブベイズでは$\boldsymbol{d}$をn枚の手札カードの相対強さr_iと場のカードの相対強さfの集合

$$\boldsymbol{d} = \{f, r_1, r_2, \cdots, r_i, \cdots, r_n\}$$

とし，lは提出カードの相対強さとした．ここで相対強さとは場に出ていないカードの中での強さのことを示す．よって相対強さは，ゲーム開始時には表 1 と同じになるが，ゲームが進みナンバーjのカードがすべて出るとナンバー$j-1$以下のカードの強さは 1 増える．例えば，ゲーム開始時の K の相対強さは 11 であり，ゲームが進んで A がすべて出された後，K の相対強さは 12 となる．

改良版のナイーブベイズでは場にカードがある時の着手は$L(s,l)$と表す．ここでsは着手Lの後，手札内の複数組が崩れるか，階段組が崩れるか，縛り状態であるかを表すビット列となる．複数組または階段組が崩れるとは手札内に着手のカードを使って着手より多く出す組み合わせがある時に，着手をしてその組み合わせがなくなることを示す．改良版のナイーブベイズでの場にカードがある時の着手$L(s,l)$の決定は以下のようにする．

$$L(s,l) = \underset{L}{\operatorname{argmax}}(P(l|\boldsymbol{d})P(s))$$

ここで$P(s)$はsの出現確率となる．また，場にカードがない

場合の着手$L(s,l,m,t)$と表す．ここでmは提出カードの枚数，tは提出カードの種類(複数または階段)を表す．改良版のナイーブベイズでの場にカードがない場合の着手$L(s,l,m,t)$以下のようにして決定する．

$$L(s,l,m,t) = \underset{L}{\operatorname{argmax}}(P(l|D)P(s)P(m)P(t))$$

ここで$P(m)$は場にカードがないときに枚数mの着手を行う確率，$P(t)$ 場にカードがないときに種類がtの着手を行う確率を表す．

5. 評価実験

評価実験では提案手法で示した相手モデルとベースラインとしてランダムに着手を行う相手モデルを実装した．相手モデルは 2010 年に優勝したモンテカルロ法を用いる snowl， UEC コンピュータ大貧民大会で配布されている default の着手をそれぞれ学習し，評価する．

実験用に学習対象のクライアントと default との 1 試合 1000 ラウンド(1 ラウンドはカード配布から大貧民決定までとする)の対戦での着手 5 試合分を記録しておく．実験では記録した着手を使いゲームのシミュレーションを行い，学習と評価を行う．シミュレーションではラウンド開始前，今までのラウンドでの着手を学習して相手モデルを作成する．そのあと，次のラウンドのシミュレーションを行う．シミュレーション中の相手モデルの評価ではモデル化対象の番が来たときの状態を評価対象モデルに入力し，着手を決定する．その着手と実際の着手が正しいか比較する．評価は相手モデルの正答率とし，以下のようにして求める．

$$\text{正答率} = \frac{\text{正解した回数}}{\text{可能手が 2 つ以上(パスは除く)の回数}}$$

6. 実験結果

ランダムに着手を行う相手モデルの正答率を表 2，ナイーブベイズによる相手モデルの正答率を表 3，改良版ナイーブベイズによる相手モデルの正答率を表 4 に示す．

表 2．ランダムによる相手モデルの正答率

クライアント	default	snowl
正答率	20.55%	19.98%

表 3．ナイーブベイズによる相手モデルの正答率

クライアント	default	snowl
正答率	48.52%	24.42%

表 4．改良版ナイーブベイズによる相手モデルの正答率

クライアント	default	snowl
正答率	64.52%	41.04%

実験の結果，改良版のナイーブベイズが最も正答率がよくなっており，2010 年に優勝したモンテカルロ法を用いる snowl クライアントに対し，4 割程度の近似ができたことがわかる．これは文献[6]での調査結果の snowl と default との一致率より低くなっている．しかし，相手の近似に default の着手をそのまま使用するより，相手モデルを使用するほうが良いと考える．これは，近似に default の着手をそのまま使用すると着手のあまりにも決定的になりすぎてしまい，「多様なシミュレーション」ができなくなるからである．相手モデルであれば，Softmax 法などと組み合わせることで「多様なシミュレーションと現実的なシミュレーション」のトレードオフの関係がある程度維持することがでる．また，snowl で使われる Policy Gradient Simulation Balancing との連携もとることができる．

7. まとめ

本研究ではプレイアウト中の相手の着手を決定するために使用する，ナイーブベイズと素性を改良した改良版ナイーブベイズにより相手の着手を学習して作成した相手モデルの評価を行った．実験の結果，モデル化は着手方法が決まっている default より，モンテカルロ法を使った snowl クライアントのほうが難しいことが分かった．また，改良版ナイーブベイズは通常のナイーブベイズより精度を大幅に向上させ，モンテカルロ法を用いている snowl クライアントに対して，4 割程度の近似をすることができた．

今後，さらに精度の良い，他の相手モデル作成方法の考案と評価を行う．

参考文献

1) 伊藤祥平，但馬康宏，菊井玄一郎：大貧民におけるゲーム中着手を反映させたプレイアウトによるモンテカルロ法，第 7 回 エンターテイメントと認知科学シンポジウム (2013) ．
2) 須藤郁弥，成澤和志，篠原歩：UEC コンピュータ大貧民大会向けクライアント「snowl」の開発，第 2 回 UEC コンピュータ大貧民シンポジウム(2010) ．
3) 須藤郁弥，篠原歩：モンテカルロ法を用いたコンピュータ大貧民の試行ルーチン設計，第 1 回 UEC コンピュータ大貧民シンポジウム(2009).
4) 小沼啓，本多武尊，保木邦仁，西野哲郎：コンピュータ大貧民に対する差分学習法の応用，情報処理学会研究報告．GI，[ゲーム情報学]，vol. 2012-GI-27，No. 1 (2012).
5) 綾部孝樹，大久保誠也，西野哲郎：大貧民プログラムの n-gram 統計による特徴抽出とクラスタ分析，情報処理学会研究報告．Vol. 2013-MPS-93，No. 2 (2013).
6) 吉原大夢，阿倍野なつみ，渡邊佑介，大久保誠也：提出手比較による大貧民プレイスタイル解析情報処理学会研究報告．Vol. 2012-GI-28，No. 7 (2013).