Vol.2011-MPS-86 No.31
Vol.2011-BIO-27 No.31
2011/12/2

多人数不完全情報ゲームのモンテカルロ木探索における推定の効果

西野順二[†1]　西野哲朗[†1]

不完全情報多人数ゲームのうち大貧民を含むトリック型ゲームでは冒頭の偶然手番で選ばれたカードの分配状態が明らかでなく、ゲームの進展に応じて徐々に情報が明らかになる。このようなゲームを対象としてモンテカルロ木探索によって最良な決定をしようとするとき未知情報である相手情報すなわち状態の推定は重要と考えられる。しかし実験的には状態の推定を行わないときと比較して、その効果はあまり大きな寄与がみとめられなかった。本論文ではこの事実にもとづいて、不完全情報ゲームの状態推定が、多人数ゲームのモンテカルロ木探索において果たす効果について、状態の集合を最適な着手による同値類に分けることで、定性的な分析を行った。とりうる状態数が選択可能な合法手に対して非常に大きいため、複数の状態を仮定して探索するモンテカルロ木探索にたいして、状態の確定的な推定はあまり重要ではないことを示した。

Ineffectiveness of Situation Estimation for Monte Carlo tree search on multi player Game with Imperfect Information

Junji Nishino[†1] and Tetsuro Nishino[†1]

A trick based card game is an imperfect information game that has a chance move at the beginning of the game and then the whole situation is closed for each other players. This information is partially reveal as the game goes on. A situation estimation process is thought as a very important factor to make program playing the game using Monte Carlo tree search. In this paper grouping analysis on the situation set according to similarity of derived moves from the situation group, in order to discusses the effectiveness of the situation estimation process with UCT search. As a result, we show that the mass of situations causes ineffectiveness of the situation detection and its usage.

1. はじめに

本論文は、展開型の不完全情報多人数ゲームの着手決定に、モンテカルロ木探索を用いるときの不完全情報の推定効果について検討する。とくに、カードゲームや麻雀などゲーム開始時のシャッフルによる一度の偶然手番をもつ不完全情報ゲームを考える。

囲碁など大きな探索空間を持つゲームにおいてモンテカルロ木探索が高い効果をあげている[1)]。不完全情報ゲームにおいてもモンテカルロ木探索が有効であることが分かってきているが、その理論的な説明はまだ明らかではない。最近になって完全情報を仮定したモンテカルロ木探索 PIMC(Perfect Information Monte Carlo) の不完全情報ゲームへの応用について、その性質の検討をするための対象ゲームを分類する指標の提案などが行われるようになった[2)]。

この文献の中で不完全情報ゲームをゲームの進展にしたがって捨て札などの情報が増えることで状態の未知情報の推定ができるトリック型ゲームと、ポーカーのように最後まで明示的な情報が増えないものに分類している。トリック型ゲームでは着手決定において、未知状態の推定を用いればより有効な意思決定が可能となるように思われる[3)]。しかしながら実験的には必ずしも有効と言えないゲームもあることが報告されている[4)]。

本論文では、状態推定があまり有効とならないようなゲームの性質について、多人数ゲーム木の探索の組み合わせモデルにもとづき、推定の効果が限定的となることを示し、コンピュータプレイヤの開発に資することを目的とする。以下では具体例として日本で広く知られているトリック型のカードゲームであり、コンピュータ大会が開催されている大貧民をとりあげる。

2. 多人数ゲーム

多人数ゲームは、ゲーム木の利得として線形に比較不能なベクトル値を取る。コンピュータ大貧民[5),6)] や麻雀など早い者勝ちで展開型のゲームの場合は各プレイヤの順位点がこれにあたり、プレイヤ人数の要素を持つベクトルである。一対比較できないため、合理的な方法では着手を一つに決めることのできないノードが多数存在し、二人ゲームで有効であった min-max 探索が安定的にできないという特性をもつ。min-max 法の自然な拡張である

†1 電気通信大学
University of Electro Communications

Vol.2011-MPS-86 No.31
Vol.2011-BIO-27 No.31
2011/12/2

Max^n 法では、どのノードにおいても均衡であるような手を見つけることができ、ある程度の強さを持つ。しかし複数の着手候補の一つでしかなく、探索木全体での最適性は保証されないため、他の手法で作られたプレイヤの方がより強いことが多い。

最近では、囲碁など大規模な完全情報二人ゲームで有効であったモンテカルロ木探索がさかんに応用され、高い効果をあげている[7]。モンテカルロ木探索は多腕バンディット問題を効果的に解くアルゴリズム UCB を二段以上の探索に応用したアルゴリズム UCT にもとづく[8]。モンテカルロ木探索を行うと、多人数性のため決定できない複数の選択をまんべんなく探索することができ、ノードの評価値としてはその平均を取ることになってこれは混合戦略となるとされている[7]。

3. 不完全情報ゲーム

不完全情報ゲームは、麻雀やカードゲームなどプレイヤごとに得られる状態の情報が部分的で不完全なゲームである。ゲームの状態が分からないとモンテカルロ木探索を含む探索的な手法は適用できない。そこで、可能な状態集合からランダムに状態を仮定しそのうえで、完全情報ゲームと同様にモンテカルロ木探索を実行することが行われている。このような仮定をモンテカルロサンプリングと呼ぶ。

Long らは不完全情報ゲームを大きく二種類に分けている[2]。一つはトリック型ゲームで大貧民もこのタイプであり、ターンごとにカードを見せ合うことでゲームが進み徐々に情報が明らかになる。二つ目はポーカー型ゲームで勝敗決定までのプロセスでは明らかな情報開示のないゲームである。ポーカー型ゲームについては Counter-factual Regret algorithm(CFR)[9] により状態数を削減することで有効なプレイが行われている。麻雀も情報開示が進む意味では大貧民と同じくトリック型ゲームに近いが、非開示の手札が 13 あり、見えない山札も平均的に 50 以上と多いためポーカーに近い面ももっている。このような両者の中間的なゲームともいえる麻雀においても、モンテカルロ木探索で良好な結果が得られている[10]。

4. 状態推定と多人数ゲーム木の探索の組み合わせモデル

大貧民を含むトリック型の展開型ゲームでは、冒頭の偶然手番でカードや牌がシャッフルおよび配布される。麻雀では順序の決められた山札も作られる。これらの手札や山札はこれ以降のシャッフルはなく、ゲーム終了まで部分的に非開示の中で確定的に状態変化していく。

あるシャッフルによる初期配布によるゲームの状態を $s_i \in S$ とする。その種類全体は大貧民では $|S| = 53!/(10! * 10! * 11! * 11! * 11!) \approx 10^{33}$ 通りである。ゲーム中はこの状況の集合のうちの一つから進展した局面にいることになる。

相手手札やひいては状態の推定は s_i の推定と同値である。実際には完全に確定できないため、S のうち可能性のある部分集合とその要素の出現確率 $f(s_i)$ をもとめていることになる。

ゲーム中に推定結果が得られたとき、次はその状況に対する最適着手を決定する。初期状況から進んだ現在の状態をいま s_i と書くことにする。この状態が分かったならば、ゲームは完全情報となるので、モンテカルロ木探索による $search(s_i)$ による意思決定をおこなうことができる。この局面での合法手の集合にたいし意思決定により各候補手に対する評価値ベクトル $v(j \mid s_i)$ が求まり、そのうちから着目するプレイヤに対応する要素を最大とする候補手 j が選ばれる。

多くの不完全情報ゲームに対する UCT アルゴリズムでは[7),10)]、状態推定は行わず、s_i の生成をモンテカルロサンプリングで行い、一つのプレイアウトを生成しまたサンプリングしなおしている。いっぽう状態を推定するモデルの方が、情報の分類ができより良い着手決定に役立つ可能性があり、毎回ランダムに状態の仮定を行うのは非効率とも考えられる。

実際のところ大貧民など偶然手番で取りうる状態の多いカードゲームなどでは、最適着手は状態 s_i に依存して変化するとはかぎらない。推定が完全にできたとすればそのたったひとつの s_i についてのみモンテカルロ木探索をおこなえばよい。しかし不明であってもある一定の範囲内で仮定した状態のいくつかでモンテカルロ木探索を行った結果が、同じ着手をさすことになるならば、そもそも s_i を一つに特定する必要がない。このような観点にたち、以下では対象とする五人大貧民での定性的な分析を試みる。

4.1 偶然手番による状態集合

五人大貧民でのカードの配り方は $53!/(10!^2 \times 11!^3)$ あり、開始時点ではそのうちのどれであるか分からない。ある時点までゲームが進むことで情報の不完全性が減ることは、開示された履歴 h から与えられる制約によって可能な状態の集合 $\{S_i\}$ が小さくなることである。

n 人のカードゲームでの状態 S は場や山札を含む複数のプレイヤへの配分と分割で定義でき $s = (t_1, t_2, \cdots p_1, p_2, \cdots p_n)$ の組で表せる。ここで各プレイヤの手 $p_i = \{c_j^i\}$ カード c の集合であり、場や山札は $t_k = (c_1^k, c_2^k \cdots c_{m_k}^k)$ という順序づけられた m_k 枚からなるカードの組である。

4.2 状態の推定

不完全情報ゲームのうちトリック型ゲームのように徐々に情報が開示されるゲームの場合、この履歴を利用して不完全情報の推定をおこない、なるべく完全情報に近づけることが

Vol.2011-MPS-86 No.31
Vol.2011-BIO-27 No.31
2011/12/2

有利であると考えられている。ゲームの履歴 h から状態の推定を行うことは、初期の偶然手番で生成された状態の集合 S から必然的に除外されるカードを含まない状態からなる部分集合 $S_h \subseteq S$ をもとめることと等しい。推定された状態の集合 S_h の要素が一つしかないとき、完全情報ゲームとなる。たとえばゲーム局面が進み二人しか残っていない場面では、完全記憶をもっておけば相手の手はすべて分かることになる。

ゲームの性質と相手モデルによって、ある状態についての存在可能性を与えられることがある。たとえば大貧民で出せるカードが複数あるときに、あえて強いカードを先に出すことは戦略的にまれであるため、逆により弱いカードをもっていない可能性が高まる。須藤らの相手手札推定モデル[11] はこの仮定にもとづき、ELO レーティングと囲碁のパターン学習にならった MM アルゴリズムと Bradley-Terry 推定モデル[12] を用いて、過去の条件付き出現率を利用した確率的な推定を行っている。

一般にある履歴 h 後の場面では、各状態の存在可能性分布を与える $f_h(s), s \in S_h$ を考えることができ、$f_h(s_i) = 0$ は s_i が確定的に除外される状態であることを表す。なお $\sum f_h(\cdot) = 1$ である。

4.3 状態に対応した最適手のモンテカルロ木探索

あるひとつの状態 s_i が与えられれば完全情報ゲームとしての着手決定を行うことができる。探索木が十分に小さければ全域探索をし、それができないほど大きい場合には、モンテカルロ木探索をおこなえばよい。モンテカルロ木探索によって得られるのは n 個の各子枝 j に対するプレイアウトを総合した式 (1) に示した評価値ベクトルの集合となる。

$$\{v_1, v_2, \cdots v_n\} = search(s_i) \tag{1}$$

ここで v_j をとってみると、g_k を m 人のプレイヤのうちプレイヤ k が得る利得として次の (2) 式となる。

$$v_j = (g_1^j, g_2^j, \cdots g_m^j) \tag{2}$$

ルートのプレイヤを k とすると、探索木全体の結果としてプレイヤ k にとって n 個の枝のうち、最大の利得 g_k を与える手を選択すればよい。

$$j = argmax_{(1 \le j \le n)} g_k^j \tag{3}$$

4.4 推定とモンテカルロ木探索による最適手の選択

従来の方法では、モンテカルロ木探索の 1 プレイアウトごとに状態を設定し、それらの総合によって UCT 探索を行いルートノードの各枝への評価値づけをおこなっている。毎回の状態が違うため 2 段目以降の他プレイヤの手番ノードはプレイアウトごとに大きく異なる展開が多くなる。

すなわち、ある一つの状態 s_i に対する評価値ベクトルの集合 $\{v_1, v_2, \cdots v_n\}$ のうち、一つの枝 j だけ $v_j^{s_i}$ としてもとめたことになる。計算回数が十分であれば同じ状態 s_i に対する他の枝もプレイアウトで調べることになるが、通常の探索の打ち切りを行う状況ではその期待は小さい。

状態の可能性分布を $f(s)$ とし、十分な数のプレイアウトが実施されたとすると、その総合結果の評価値ベクトルの期待値 v_j は次のようになる。ただし $\sum f(\cdot) = 1$ である。

$$v_j = \sum_{s_i \in S} f(s_i) v_j^{s_i} \tag{4}$$

4.5 推定の効果

不完全情報の推定は状態の実現可能性を履歴で条件づけられ関数 $f_h(s)$ で表すことである。

あらためて式 (4) の右辺を考えると、状態集合 S は、式 (5) に示したようにある同じ評価値ベクトル v^i をとる関係によって同値類に分割することができる。

$$s_i \ \sim s_k \ s.t. \ v_{s_i} = v_{s_k} = v^i \tag{5}$$

v^i による S の同値類を S_i と表す。これにより式 (4) は次のように書き換えることができる。

$$v_j = \sum_{all S_i} \{ v_j^i \sum_{s_i \in S_i} f(s_i) \} \tag{6}$$

状況が決まればモンテカルロ木探索による評価値ベクトルが定まる。この関係は一対一ではなく、その評価値ベクトルを導くある手 j が選ばれるためにはある特定の状況 s_i だけでなく、その同値類のどれか一つであればよい。

最大の利得 g_k を与える評価値ベクトル $v_j = (g_1, g_2 \cdots g_n)$ からの式 (3) による手 j の選択過程でも、g_k^j が他の v_l の g_k^l と比べて大きければよく状況 s_i との関係も 1 対 1 ではない。

このため推定された状況も得られた評価値ベクトルも、正解とは異なるが、着手は同じであることがしばしば発生する。これは不完全情報である状態集合の大きさ $|S|$ が可能着手・合法手の数 $|m|$ とくらべて非常に大きいため、同値類の大きさも大きくなりやすいからである。そこで、厳密な意味での適合性能の悪い推定器をもちいたしとても、それが同値類に含まれる状態を示す割合が高ければ、結果として正解着手を選ぶことができたことになる。

いっぽう、原理的に厳密な推定が必要となるのは、ある履歴状態において正解手を導く状態の同値類が極端に小さいときである。たとえば正解の状態 s_i 以外の状態集合ではどの $s \in S$ をとっても正解が得られないことがある。このときは、正解よりも多くの誤った状態を推定結果として残してしまうと、評価値ベクトル中の値の期待値は誤った状態によって小

Vol.2011-MPS-86 No.31
Vol.2011-BIO-27 No.31
2011/12/2

さくなってしまう。そもそもこのような特異性の高い探索はモンテカルロ木探索には向いていないといえる。

逆に状態の推定と最適手の決定の関係は緩やかなものであり、とりうる状態が多い対象ほどその厳密な推定の必要性は少ない可能性が高い。つまり正解状態を厳密に推定するよりも、状態推定としては誤っているとしても正解を与える状態を多数生成できる推定器のほうが強い着手決定には役立つ。

4.6 大貧民における推定の効果

実対象の例として、トランプゲーム大貧民に対するコンピュータプレイヤを用いて行われた推定効果の実験[4] では、次のような結果が得られている。推定モデルを停止してもパフォーマンスが大きく変化しない。特定のコンピュータプレイヤに適応学習した推定モデルが、ことなるアルゴリズムのプレイヤに対しても有効であった。

この大貧民では五人のプレイヤに 10 ないし 11 枚の手札を配り、単調に減っていくため、平均的な分岐因子は 2 から 5 手程度である。半分の枚数をプレイし終わった全体の残り枚数が 25 枚程度のばあい可能な状態の全数の最大値は $25!/5!^5 = 10^{14}$ 程度であるから、一人 5 枚のうちどれか 1 枚が最適手であることを支持する状況、同値類の要素の数は 10^{13} を超えることになる。

多人数で行われることで、不完全情報の源である、手札の組み合わせによる状態の数が増えることも明らかである。たとえば、53 枚を 53 人に配る場合の数は 53! となり、5 人に配るよりも大きい。

このことから、先行論文で示された大貧民では本論文で指摘した状態と最適手のゆるやかな関係が形成され、最終的な着手決定では有効に働いたものと考えられる。

5. ま と め

本論文では、大貧民のような多人数トリック型のゲームでは、合法手の数が少なくかつ状態の可能性が極端に多いという、最適手の数と状態数との対応関係から、厳密な状態推定の有効性が高くないことを示した。

トリック型の不完全情報多人数ゲームへのモンテカルロ木探索の適用において、状態推定をする上での方策は次のように考えられる。

厳密な状態推定の効果は限定的であるため、推定器にかける計算時間をより多い種類の状態の生成とモンテカルロ木探索にかけた方が良いかもしれない。とくに推定器の開発と調整にかけるコストは慎重になる必要がある。

逆に精密な一つの正解状態の推定ではなく、ある局面での最適な着手を与えることのできる状態の類が分かればよい。このためには着手そのものを求めずに同じ着手を与えるような状態の類を探す方法が必要である。ポーカーの状態縮約で提案された[9] 探索空間の実態を減らすようなアルゴリズムによって、着手について同相な状態の発見が今後の課題である。

参 考 文 献

1) 美添一樹：モンテカルロ木探索-コンピュータ囲碁に革命を起こした新手法，情報処理，Vol.49, No.6, pp.686–693 (2008).
2) Long, J., Sturtevant, N.R., Buro, M. and Furtak, T.: Understanding the Success of Perfect Information Monte Carlo Sampling in Game Tree Search, *Proceedings of the 24th. AAAI Conf.*, AAAI, pp.134 – 140 (2010).
3) 小田和友仁，上原貴夫：コンピュータブリッジにおける他者のモデルを考慮したゲーム木探索の提案 (知識処理)，情報処理学会論文誌， Vol.47, No.11, pp.3005–3016 (2006).
4) 西野順二，西野哲朗：大貧民における相手手札推定，情報処理学会研究報告 2011-MPS-85，No.9 (2011).
5) 西野哲朗：第 1 回 UEC コンピュータ大貧民大会 (UECda-2006) の実施報告，情報処理学会誌， Vol.48, No.8, pp.884–888 (2007).
6) 大久保， 小林， 本多， 眞鍋， 青木， 柿下， 小松原， 西野：第 1 回コンピュータ大貧民大会 (UECda-2006) の報告，情報処理学会ゲーム情報学研究報告，Vol.GI-17, pp.25–32 (2007).
7) Sturtevant, N.: An Analysis of UCT in Multi-player Games, *Computers and Games*, Lecture Notes in Computer Science, Vol.5131, Springer, pp.37–49 (2008).
8) Kocsis, L. and Szepesvari, C.: Bandit based Monte-Carlo Planning, *the 17th European Conf. on Machine Learning*, pp.282 – 293 (2006).
9) Zinkevich, M., Johanson, M., Bowling, M. and Piccione, C.: Regret Minimization in Games with Incomplete Information, *Neural Information Processing Systems*, No.20, pp.1729 – 1736 (2008).
10) 三木理斗，三輪 誠，近山 隆：UCT 探索による不完全情報下の行動決定，第 14 回ゲームプログラミングワークショップ 2009，pp.43– 50 (2009).
11) 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講演予稿集，電気通信大学 (2010).
12) Humter, D.R.: MM algorithms for generalized Bradley-Terry models, *The Annals of Statistics*, Vol.32, No.1, pp.384–406 (2004).