デジタルカーリングにおけるMCTS候補手削減のための価値考慮型クラスタリングの適用

仲 亜斗夢[1,a)]　横山 大作[1,b)]

概要：デジタルカーリングは行動が連続値で投球に外乱が加わるため，モンテカルロ木探索（MCTS）を用いて有限個の候補手を生成して探索するが，候補手が多いと限られた計算量では各候補手の評価が不十分になる．われわれは先行研究で，候補手を打った後の盤面（結果盤面）をクラスタリングし，各クラスタの中心的な手だけを探索する手法を提案したが，代表手の選び方は手の価値を考慮していなかった．本研究では，少数のロールアウトで推定した得点期待値を用いて，価値の高いクラスタとその中で最も有望な手を選ぶ ClusterValue を提案し，木の中で手を外乱ありで打ち直す外乱込み評価と組み合わせる．エンドごとに抽出した 500 局面で物理シミュレーションの回数を揃えて比較した結果，外乱込みで求めた最良手との得点期待値の差（regret）は既存手法より 22%小さく，候補を絞らない全探索も上回り，その差はプレイアウト回数が少ないほど大きいことを示した．また，結果盤面のクラスタが相手の中心の石を取り除くような他と似ていない最良手や，即時の得点は低いがエンド全体では同等に有利な別の手を，独立したまとまりとして表すことを盤面の図で示した．

Action Space Reduction in MCTS by Value-aware Clustering of Stone Positions in Digital Curling

Atomu Naka[1,a)]　Daisaku Yokoyama[1,b)]

Abstract: In digital curling, actions are continuous and shots are unstable by execution noise, so Monte Carlo tree search (MCTS) searches over a finite set of generated candidate shots; with many candidates, each one is evaluated insufficiently under a limited computational budget. Our previous work clustered candidates by the boards they produce and searched only the most central shot of each cluster, ignoring shot values. We propose ClusterValue, which uses expected scores estimated by a few rollouts to keep high-value clusters and the most promising shot in each, and combine it with noise-aware evaluation that re-executes shots with noise inside the tree. On 500 positions with an equal number of physics simulations, ClusterValue reduced the regret (the expected-score gap from the best shot under noise) by 22% compared with the previous method and also outperformed exhaustive search over all candidates, with larger gains at fewer playouts. Board examples also show that the clusters separate a distinctive best shot, such as removing the opponent's center stone, and alternative shots that score less immediately but are equally advantageous over the end.

1. はじめに

カーリングは，氷上で石（ストーン）を滑らせ，ハウスと呼ばれる同心円の的の中心（ティー）にどれだけ近づけられるかを競う競技である．「氷上のチェス」と呼ばれるように，数手先の石の配置を見越した読みが要求される．デジタルカーリング [1] は，この競技を 2 次元の物理シミュレータ上で再現した対戦環境であり，モンテカルロ木探索（MCTS）を中心とした思考エンジンが研究されてきた [2], [3], [4].

カーリングの試合は複数のエンドからなり，各エンドでは 2 チームが交互に 8 投ずつ，計 16 投を投げる．エンドの終わりに，ティーに最も近い石（No.1 ストーン）を持つ

[1] 明治大学
Meiji University
[a)] ce265004@meiji.ac.jp
[b)] dyokoyama@meiji.ac.jp

© 2026 Information Processing Society of Japan

チームが，相手チームの最もティーに近い石より内側にある自チームの石の数だけ得点する．得点に数えるのはハウスの中の石だけである．エンドの最終投を投げる後攻は，最後に石の配置を変えられるため有利である．主なショットには，狙った位置に石を止めるドロー，他の石に当てて動かすヒット，相手の石に密着させて止めるフリーズ，ハウスの手前に石を置いて後ろの石を守るガードがある．ガードのうち，ハウスの中の自分の石の前に置くものをポストガード，ハウスの手前の決まった位置に置くものをプレガードと呼ぶ．また，自分の石に当ててティーの方へ押し込むヒットをレイズと呼ぶ．デジタルカーリングでは，プレイヤは 1 投ごとに石の初速 (v_x, v_y) と回転方向（時計回りか反時計回り）を選ぶ．行動は連続値であり，さらに初速には正規分布に従う誤差（以下，外乱）が加わるため，狙いどおりに投げても結果は確率的に変わる．

連続行動空間で MCTS を用いる際には，ドロー，ヒット，ガードなどの候補手を有限個生成し，それを木の分岐とすることが多い．しかし候補手を十分に用意すると分枝数が大きくなり（本論文の設定では 1 局面あたり 33〜153 手），限られたプレイアウト回数のもとでは 1 つの候補手に割り当てられるシミュレーション回数が不足する．対戦では持ち時間が決まっているため，少ない計算量で良い手を選べることは実用上も重要である．大きな行動空間や連続行動空間での MCTS については，ボロノイ分割に基づいて行動をサンプリングする VOOT [5] や，学習した遷移モデルから状態ごとに不要な行動を除く状態条件付き行動抽象化 [6] などが提案されている（2 章）．

われわれの先行研究 [7] では，各候補手を打った後の盤面（以下，結果盤面）に階層的クラスタリングを適用し，各クラスタの代表手のみを探索対象に残すことで候補手を削減する手法を提案した．結果盤面でクラスタリングするのは，外乱のあるカーリングでは「どこを狙ったか」よりも「その手がどのような盤面を引き起こすか」によって手の役割が決まると考えたためである．本論文では，この手法を MCTS に拡張したものを ClusterMedoid と呼び，既存手法として扱う．ClusterMedoid は，各クラスタの代表手として盤面上で最も中心的な候補（medoid）を選ぶが，この選び方は手の得点期待値を考慮していないため，代表手が有望な手になるとは限らない．また先行研究では，全候補を探索する手法が選んだ手との一致率で性能を評価していたが，プレイアウト回数が限られた全探索の選択は正解の代わりにならないという問題もあった．

そこで本論文では，結果盤面によるクラスタは保ったまま，クラスタの採用と代表手の選出に得点期待値を用いる手法 ClusterValue を提案する．ClusterValue は，各候補手の得点期待値を少数のロールアウトで安価に推定し，その平均値（クラスタ価値）が高い上位のクラスタから，それぞれ推定値が最大の候補を代表手として探索木に残す．さらに，木の中で手を評価するたびに外乱ありで打ち直す外乱込み評価を組み合わせ，「狙った結果」ではなく「その手を試みる価値」を評価する．

性能は各候補手を外乱ありで 200 回ロールアウトして求めた得点期待値（以下審判値）を正解の基準とし，選んだ手と最良手との審判値の差（regret）で評価した．層化抽出した 500 局面で，シミュレーションの回数を揃えた条件で比較した結果，提案手法は既存手法より regret が 22%小さく，候補を絞らない全探索も有意に上回った．その利得はプレイアウト回数が少ないほど大きかった．

さらに，結果盤面のクラスタが人間にとって意味のある単位になっている場合があることを，いくつかの盤面において確認した．クラスタは行動した後の結果の種類でまとまっており，相手の中心の石を取り除くような他と似ていない最良手や，そのエンドの得点は低いがエンド全体では同等に有利な別の手を，それぞれ独立したまとまりとして示せているような成功例が確認された．また，結果盤面による分割は，VOOT のような行動空間の分割よりも価値のまとまった領域を作ることも明らかになった．

本論文の構成は次のとおりである．2 章で関連研究を述べ，3 章で既存手法と提案手法を説明する．4 章で性能の比較実験を，5 章で既存手法との差とクラスタの視覚的評価についての考察を述べ，6 章で今後の課題を述べる．

2. 関連研究

2.1 大規模な連続行動空間での MCTS

Yee ら [4] の KR-UCT は，カーリングのシミュレータを対象に，実行時の誤差のモデルをカーネルとするカーネル回帰によって，ある行動を試した結果を，行動のパラメータが近い他の行動の評価にも用いる．KR-UCT が行動のパラメータの近さで情報を共有するのに対し，本論文は結果盤面の近さで候補手をまとめる．そのため，回転方向が逆で行動としては離れた 2 つの手でも，同じ結果盤面を生めば同じクラスタに入る．

Kim ら [5] の VOOT は，決定的な遷移を持つ連続行動空間の問題を対象に，木の各ノードでボロノイ分割に基づくブラックボックス最適化 VOO によって行動をサンプリングする．VOO は評価済みの行動を母点として行動空間をボロノイ分割し，一定の確率で現在の最良の行動のセルの中から次の行動を選ぶ．本論文の手法も候補を領域に分けて各領域を 1 つの代表で扱い，最良の候補を含む領域を必ず探索する点で VOO と共通する．一方で，領域を行動空間ではなく結果盤面の近さで作る点，外乱のある環境を扱う点，行動を新たに生成せず有限個の候補から選別する点が異なる．この対応関係は 5 章で検討する．

Kwak ら [6] は，行動が複数の部分行動の組み合わせからなり，遷移が決定的な問題を対象に，状態条件付き行動抽象化を提案した．この手法は，学習した潜在ダイナミク

© 2026 Information Processing Society of Japan

スモデルから現在の状態の遷移に関係する部分行動を推定し，木の各ノードで関係しない部分行動の探索を省く．本論文の手法も，局面ごとに行動の結果に基づいて候補をまとめる点で，状態条件付きの行動抽象化の一種とみなせる．ただし本論文は，学習したモデルではなくシミュレータで得た結果盤面を直接用いるため学習を必要とせず，遷移に関係しない行動を除くのではなく，結果の近い行動を 1 つのまとまりとして扱う．

2.2 先行研究の課題と本研究の位置づけ

デジタルカーリングにおける探索では，連続な行動空間から候補手を有限個に絞ったうえで探索する方法が主に用いられてきた．大渡ら [2] の「歩」は，ドロー，ヒット，ガードなどのショットの種類ごとに盤面に応じた候補手を生成し，有望な候補手に多くのプレイアウトを配分する MCTS である．また, 本論文の提案手法の候補手生成も歩を参考にしている．加藤ら [3] は，行動空間をグリッドで離散化し，外乱を考慮した Expectimax 探索を行った．いずれの手法でも 1 局面あたりの候補手は多く，限られた計算量で深く探索するには候補手をさらに絞る必要がある．

われわれの先行研究 [7] では，この課題に対し，各候補手の結果盤面を階層的クラスタリングでまとめ，各クラスタの代表手（medoid）だけを評価することで候補手を削減した．この方法は，学習を用いずに局面ごとに行動をまとめる点で前節の状態条件付き行動抽象化に近く，結果盤面の近さで領域を作る点で VOOT とは異なる立場をとる．一方で，先行研究に残っていた課題としては例えば, 探索は深さ 1 の原始モンテカルロ法にとどまり，候補手は外乱なしで打った結果盤面から評価していたことや, 代表手を盤面上の中心性だけで選ぶため，その代表手の価値を考慮せずに代表手の選択を行っていたことが挙げられる．

本研究は，この先行研究を発展させたものである．結果盤面によるクラスタは保ったまま，クラスタの採用と代表手の選出に得点期待値を用い，これを深さ 3 の MCTS に組み込んで外乱込みで評価し，外乱込みのロールアウトで求めた審判値に基づく regret で評価する．

3. 提案手法

本論文で扱う手法は，いずれも MCTS の各ノードで候補手生成器が N 個の候補手 $a_1, \ldots, a_N$ を生成し，それぞれを外乱なしでシミュレーションして結果盤面 $s_1, \ldots, s_N$ を求めたうえで，どの候補手を子として残すかだけが異なる．既存手法 ClusterMedoid は，結果盤面をクラスタリングし，各クラスタで盤面上最も中心的な候補を代表手として残す．提案手法 ClusterValue は同じクラスタを用いたまま，次の 3 点を変更する．一つ目，各候補手の得点期待値を少数のロールアウトで推定し，クラスタの価値を求める（3.2 節）．二つ目，代表手を medoid ではなくクラスタ内で推定値が最大の候補とし，価値の高い上位のクラスタだけを子として残す（3.3 節）．三つ目，木の中で手を評価するたびに外乱ありで打ち直す外乱込み評価を用いる（3.4 節）．既存手法との違いを**表 1** にまとめる．4 章では，これらすべての変化について検証し, 候補を絞ること自体の利得を全探索との比較する．

表 1 既存手法 ClusterMedoid と提案手法 ClusterValue の違い

	ClusterMedoid	ClusterValue
代表手の選び方	medoid（盤面距離の総和が最小）	クラスタ内で推定値 $\hat{e}_i$ が最大
採用するクラスタ	すべて	価値の上位 K 個（最良候補のクラスタを必ず含む）
ルートの子の数	$\lceil 0.2N \rceil$（平均 14.4）	$\lfloor P/50 \rfloor$
事前ロールアウト	なし	各候補 $R_{\mathrm{pre}} = 5$ 回
外乱の考慮	なし	あり

3.1 既存手法：ClusterMedoid

ClusterMedoid は，先行研究 [7] の候補削減を MCTS の各ノードに適用した手法である．まず，結果盤面どうしの距離 $d(s_i, s_j)$ を，次の 8 項目の重み付き和として定義する．

(1) 元の盤面にあった石の移動量の差
(2) 投げた石の停止位置の差
(3) 石の有無（場外に出たかどうか）の不一致
(4) 石が属する領域（ハウス内，ガードゾーン，それ以外）の不一致
(5) 既存の石に触れたかどうかの不一致
(6) 投げた石と最も近い石との距離の差
(7) その時点で数えた得点の差
(8) No.1 ストーンを持つチームの不一致

この距離に基づき，平均連結法による階層的クラスタリングで候補手を $K_c = \lceil rN \rceil$ 個のクラスタに分割する（$r = 0.2$）．各クラスタからはクラスタ内の他の候補との距離の総和が最小となる候補（medoid）を代表手とし，すべてのクラスタの代表手を子とする．ClusterMedoid の代表手の選び方は，盤面の中心性のみに基づいており，手の得点期待値を考慮していない．そのため，同じクラスタに得点期待値の高い候補と低い候補が混在する場合や，得点期待値の低いクラスタが多数を占める場合には，有望でない手に探索時間が割かれる．先行研究 [7] でも，代表点の選び方の改善と，大きなクラスタからの代表の選び方が課題として残されていた．

3.2 得点期待値の推定とクラスタ価値

ClusterValue はまず各候補手 a_i について，その結果盤面からエンドの終わりまでのロールアウトを R_{pre} 回行い，

© 2026 Information Processing Society of Japan

エンドの純得点の平均 $\hat{e}_i$ を求める．次に，ClusterMedoid と同じ手順で候補手を K_c 個のクラスタ $C_1, \ldots, C_{K_c}$ に分割し，各クラスタの価値（クラスタ価値）をメンバーの推定値の平均

$$v(C_c) = \frac{1}{|C_c|} \sum_{i \in C_c} \hat{e}_i \tag{1}$$

とする．R_{pre} は小さい（本論文では 5 回）ため，個々の候補の推定値はばらつきが大きい．そのため, クラスタ単位で平均を取ることでばらつきを抑え, どの種類の結果が有望かを比較できる．

3.3 代表手とクラスタの選び方

ClusterValue はクラスタをクラスタ価値 $v(C_C)$ を降順に並べ，上位 K 個のクラスタから，それぞれ得点期待値 $\hat{e}_i$ が最大の候補を代表手として子にする．ただし，全候補の中で $\hat{e}_i$ が最大の候補が属するクラスタは，価値の順位によらず必ず採用する．クラスタ価値は平均であるため，有望な手が価値の低い手と同じクラスタにあるとそのクラスタの価値は低くなるが，この規則によって推定値が最も高い手は必ず子に残る．子の数 K は，子 1 つあたりに確保したい訪問回数 V を用いて $K = \max(1, \lfloor P/V \rfloor)$ とする．本論文では $V = 50$ とし，$P = 200$ のとき $K = 4$ である．既存手法がすべてのクラスタの代表を子とするのに対し，提案手法は子を少数に絞り，浮いた計算量を 1 つの子あたりの訪問回数の増加に割り当てる．

3.4 外乱込み評価

通常の MCTS の木は, 外乱なしでシミュレーションした結果盤面を子の状態とする．この決定的評価では狙いどおりに投げられた仮定の下での価値で評価される．一方，実際の対戦では手は外乱を伴って実行されるため，わずかなずれで結果が大きく変わる手は，決定的評価では考慮することが難しい．本論文で正解の基準とする審判値も，外乱ありで打ち直した結果の平均と定義する．

外乱込み評価では，木の構造（候補手の生成，クラスタリング，候補の選別）は外乱なしの結果盤面から作ったまま，価値の評価だけを外乱込みで行う．具体的には，プレイアウトのたびに選んだ手をその時点で到達している実際の盤面から外乱ありでシミュレーションし直し，得られた盤面を次のノードへ運ぶ．葉のロールアウトも，この外乱込みで到達した盤面から行う．ClusterValue の事前推定でも，候補手を毎回外乱ありで打ち直してからロールアウトする．これにより，各ノードの価値は, その手順を外乱のもとで試みたときの得点期待値の推定となる．

4. 実験

4.1 実験の目的

本実験の目的は，提案手法 ClusterValue の性能を，既存手法 ClusterMedoid および候補を絞らない全探索と，プレイアウト回数を揃えた条件で比較することである．特に次の 3 点を明らかにする．

(1) 外乱込み評価を加えると，各手法の性能はどう変わるか
(2) 全探索と比べて，どのようなプレイアウト回数のときに候補を絞る方が優れるか
(3) 代表手の選び方の違いは，性能にどう影響するか

また，本実験で得たクラスタは 5.2 節にも用いる．

4.2 比較手法

比較する手法は次の 5 つである．

- AllGrid：全候補を子とする全探索.
- ClusterMedoid：既存手法.
- ClusterValue：提案手法.
- RandomK：ClusterMedoid と同じ数の候補を無作為に選んで子とする手法.
- RandomK4：ClusterValue と同じ 4 個の候補を無作為に選んで子とする手法.

RandomK は medoid による選び方を，RandomK4 は得点期待値による選び方を，それぞれ同じ子の数の無作為な選び方と比べるための対照である．各手法を決定的評価と外乱込み評価の両方で実行した．ただし RandomK4 は外乱込み評価のみである．

4.3 評価指標

正解の基準として，各局面の全候補手について，候補手を外乱ありで打ち直してからエンドの終わりまでロールアウトすることを 200 回繰り返し，手番側から見たエンドの純得点の平均を求めた．これを候補手 a_i の審判値 $q(a_i)$ と呼ぶ．審判値は「その手を試みる価値」の推定であり，探索木を用いないため各手法の実装に依存しない．手法が選んだ手 a の regret を $q^* - q(a)$ とする．ここで q^* は局面内の審判値の最大値であり，regret は小さいほど良い．審判値の標準誤差は約 0.10 点であるが，1 位と 2 位の候補の審判値の差は中央値で 0.065 点とこれより小さく，審判値だけでは判断するのが難しい．そこで, 最良手との一致率ではなく，値の差である regret で評価する．

regret の大きさの目安として，1 局面の候補手の regret を**図 1** に, 500 局面の全候補の regret の分布を**図 2** に示す．この局面では 61 手の regret が 0 から 1.40 まで広がっている．赤線は後述する ClusterValue の平均 regret（0.309）であり，この局面ではこれより小さい regret の候補は 8 手だけである．500 局面全体でも regret が 0.309 以内に入る候補は局面ごとの中央値で 14%（9 手）にとどまり，無作為に 1 手を選んだ場合の regret は平均 0.78 である．後述する各手法の regret は，この分布の左端にあたる．

さらに regret を，選別で最良手を落としたことによる損

© 2026 Information Processing Society of Japan

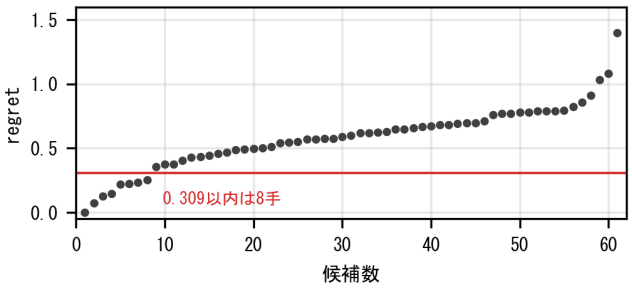


図 1　1 局面の 61 手の候補を regret の小さい順に並べた例

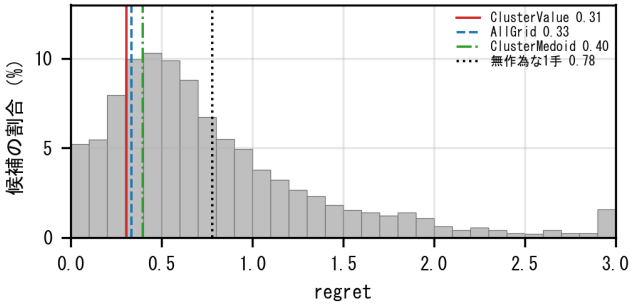


図 2　500 局面の全候補の regret のヒストグラム（各局面の重みを等しくし，3 以上は右端の区間に含めた）

失 (選別損失) と，残りの損失 (木の損失) の和に分けて分析する．ルートの子として残った候補の集合を $\mathcal{S}$ とし，その中で最も大きい審判値を $q_{\mathrm{c}} = \max_{a_i \in \mathcal{S}} q(a_i)$ とすると，

$$q^* - q(a) = (q^* - q_{\mathrm{c}}) + (q_{\mathrm{c}} - q(a)) \qquad (2)$$

であり，右辺の第 1 項が選別損失，第 2 項が木の損失である．手法は子の中から手を選ぶため，どちらの項も 0 以上である．選別損失は最良手が子に残っていれば 0 となり，木の損失は子の中で最良の手を選べれば 0 となる．

4.4　実験条件

実験条件を**表 2** に示す．評価局面は，歩の方策関数 [2] を移植した方策どうしの自己対戦（序盤の数投は無作為に選択）から抽出した．先頭の 200 局面は主に石の数と進行段階で層化して抽出し，残りの 300 局面は，盤面の型（石が 0〜1 個，2〜4 個，相手の No.1 ストーンを除去できる形，密着できる形，空きが多い形，石が 6 個以上）と進行段階で層化して抽出した．このため，本論文で示す局面の割合は実戦での出現頻度とは異なる．

探索木は深さ 3 まで展開し，エンドの最終投を越える枝は終端としてそのエンドの得点を報酬とする．ロールアウト方策は，各手番でハウスを覆う位置へのドロー，相手の No.1 ストーンへのヒットなど 6〜10 個の候補を生成し，確率 0.3 で無作為に，それ以外は 1 投後の盤面で数えた得点が最大となる候補を選ぶ．ロールアウトの投球は外乱ありでシミュレーションを行う．ただし，本論文の比較はシミュレーション回数を揃えたものであり，実時間の比較ではない．

表 2　実験条件

項目	設定
評価局面	自己対戦から層化抽出した 500 局面
候補手の数 N	1 局面あたり 33〜153 手（平均 70.6）
シミュレータ	デジタルカーリング [1]
外乱	初速の大きさと向きに正規分布の誤差（標準偏差 0.0076 m/s，0.0018 rad）
探索の深さ	3
プレイアウト回数 P	200
葉のロールアウト R	10 回
子の選択	UCB1（各ノードの手番側の価値を最大化）
報酬	探索開始時の手番側から見たエンドの純得点
最終的に選ぶ手	ルートの子のうち訪問回数が最大の手
クラスタ数 K_c	$\lceil 0.2N \rceil$
事前推定 R_{pre}	5 回（ClusterValue のみ）
ClusterValue の子の数 K	4（$V = 50$）
乱数のシード	5 通り

表 3　各手法の regret (500 局面× 5 シードの平均，$P = 200$．小さいほど良い)

手法	決定的評価	外乱込み評価
AllGrid（全探索）	0.418	0.334
ClusterMedoid（既存手法）	0.488	0.396
RandomK（$\lceil 0.2N \rceil$ 個）	0.491	0.383
RandomK4（4 個）	–	0.552
ClusterValue（提案手法）	0.423	0.309

4.5　実験結果

4.5.1　全体の比較

表 3 に各手法の regret を示す．外乱込み評価の条件では，提案手法 ClusterValue の regret が 0.309 と最も小さかった．これは同じプレイアウト回数の全探索 AllGrid より小さく（$p = 2 \times 10^{-4}$），既存手法 ClusterMedoid（0.396）より 22%小さい（$p = 4 \times 10^{-8}$）．外乱込み評価の条件で見た局面単位では，ClusterValue の方が regret が小さかった局面は，AllGrid に対して 297，ClusterMedoid に対して 282 であり，大きかった局面はそれぞれ 198 と 205 であった．決定的評価の条件でも，ClusterValue は ClusterMedoid より有意に小さく（0.423 と 0.488，$p = 6 \times 10^{-4}$），AllGrid（0.418）とは差がなかった（$p = 1.0$）．

4.5.2　外乱込み評価の効果

外乱込み評価は，候補の選び方によらず regret を下げた（表 3 の左右の列）．regret の減少は，AllGrid で 0.084（$p = 3 \times 10^{-4}$），ClusterMedoid で 0.092（$p = 8 \times 10^{-8}$），ClusterValue で 0.114（$p = 9 \times 10^{-15}$）であった．ClusterValue の regret を分解すると，選別損失は 0.230 から 0.211 とほぼ変わらず，木の損失が 0.193 から 0.098 に半減していた．ここから，残した子の評価が正確になったことによっ

© 2026 Information Processing Society of Japan

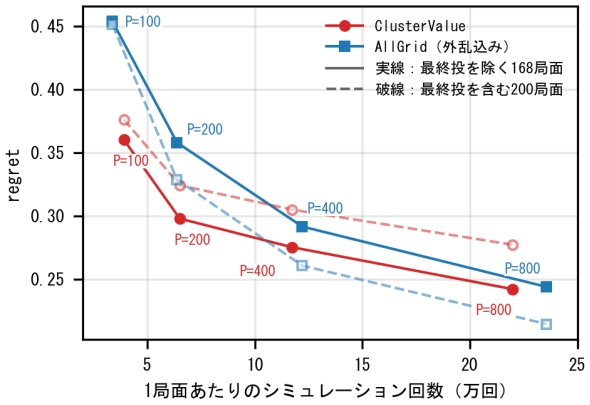


図 3 プレイアウト回数 P を変えたときの regret．実線は最終投を除く 168 局面，破線は最終投を含む 200 局面 (いずれも 5 シードの平均)．横軸は 200 局面での 1 局面あたりの物理シミュレーション回数の平均

て改善されたと推測できる．また，選んだ手の審判値の標準偏差は外乱込み評価の有無によらず 1.68 であった．外乱込み評価によって安全な手を好むようになったのではなく，外乱のもとで実際に実現する価値の高い手を選ぶようになったと解釈できる．ただし，審判値も外乱込みで求めているため，この改善の一部は，木の評価のルールを正解の基準のルールに揃えたことによると考えられる．

4.5.3 計算量と全探索との比較

$P = 200$ では，外乱込み AllGrid は全候補を子に持つため選別損失は 0 であり，regret 0.334 はすべて木の損失である．これに対して ClusterValue は，選別損失 0.211 を払う代わりに木の損失を 0.098 に抑えている．平均 70.6 個の子にプレイアウトを分散させると各子の評価が不十分になり，選別による損失を払ってでも少数の有望な子を厚く評価した方が良い手を選べることが分かる．

この AllGrid に対する ClusterValue の regret の減少 (以下, 利得) がプレイアウト回数の不足によるものであれば，P を大きくすると利得は縮むと予想される．そこで 500 局面の先頭 200 局面について，P を 100, 200, 400, 800 と変えて ClusterValue と外乱込み AllGrid を比較した (**図 3**)．ClusterValue の子の数 K は P に応じて,2, 4, 8, 16 と増える．図の横軸は 1 局面あたりの物理シミュレーション回数であり, 同じ P でも手法によって異なるため, P ではなく実際の回数で並べた. 縦軸は regret であり, 下にあるほど良い. 同じ P の 2 点の縦の差が利得にあたり，同じ高さの 2 点の横の差が，同程度の regret に達するのに必要な計算量の差にあたる．最終投を除く 168 局面 (実線) では，両者の regret の差 (ClusterValue − AllGrid) は $P = 100$ の -0.094（$p = 2 \times 10^{-5}$），$P = 200$ の -0.060（$p = 3 \times 10^{-5}$）から，$P = 400$ の -0.016（$p = 0.12$），$P = 800$ の -0.002（$p = 0.17$）まで単調に縮小し，どの P においても ClusterValue の方が悪くなることはなかった（p は片側の Wilcoxon 検定）．同じ高さの点を横に比べると，ClusterValue は全探索の約 5〜6 割のシミュレーション回数で済んでいる．たとえば ClusterValue の $P = 100$（3.9 万回）の 0.360 は外乱込み AllGrid の $P = 200$（6.4 万回）の 0.358 と，$P = 200$（6.5 万回）の 0.298 は $P = 400$（12.2 万回）の 0.292 とほぼ同じ結果を示す．なお，図 3 は先頭 200 局面の結果であるため，図の $P = 200$ の値は 500 局面で求めた表 3 の値と一致しない．

表 4 最良手のクラスタの大きさ別の regret の差 (ClusterValue − 比較手法)．最終投を除く 468 局面．$^{**}p < 0.01$，$^{*}p < 0.05$ (層ごとの Wilcoxon 検定，補正なし)

		対外乱込み	
大きさ	局面数	ClusterMedoid	AllGrid
1	40	+0.007	−0.130**
2	51	+0.007	−0.117**
3〜5	79	−0.056	−0.061**
6〜10	200	−0.090**	−0.028*
11 以上	98	−0.046	+0.001

一方，エンドの最終投では得点につながる手が狭い範囲の少数の手に限られることが多く，事前推定による選別でその手を落としやすい．そのため最終投を含めた 200 局面では，$P \geq 400$ で外乱込み AllGrid の方が regret が小さくなった．500 局面全体でも，最終投の 32 局面では ClusterValue の regret が外乱込み AllGrid より 0.287 大きい．

4.5.4 代表手の選び方の影響

子の数が同じ 4 個どうしで比べると，無作為に選ぶ RandomK4 の 0.552 に対し，得点期待値で選ぶ ClusterValue は 0.309 であった（$p = 1 \times 10^{-43}$）．一方，既存手法の medoid による選び方は，同じ数を無作為に選ぶ RandomK と差がなかった（外乱込み評価で 0.396 と 0.383，決定的評価で 0.488 と 0.491，いずれも $p > 0.8$）．結果盤面のクラスタで候補を分けても，その中からどの手を残すかを価値で決めなければ性能は上がらない．

既存手法との差がどのような局面で生じるかを，最良手（審判値が最大の手）が属するクラスタの大きさで局面を分けて調べた（**表 4**）．最良手のクラスタが 6〜10 手の局面では，ClusterValue の regret が外乱込み ClusterMedoid より 0.090 小さかった．大きいクラスタほど，medoid が価値の低い手を代表に選ぶ機会が多いためである．

4.6 実験のまとめ

以上の 3 つの問いに対する結果をまとめる．まず，外乱込み評価はどの手法でも regret を下げ，その効果は主に木の損失の減少によるものであった．次に，候補を絞ることの利得はプレイアウト回数が少ないほど大きく，最終投を除けばどのプレイアウト回数でも提案手法は全探索の利得を上回った．

© 2026 Information Processing Society of Japan

ただし最終投では，選別で最良手を落とすため全探索の方が良い．さらに，得点期待値による代表手の選び方は無作為な選び方を大きく上回ったが，medoid による選び方は無作為な選び方と変わらなかった．以上から，提案手法の性能の向上は，得点期待値に基づく候補の選別と外乱込み評価によってもたらされたといえる．

5. 考察

5.1 既存手法との性能の差

既存手法 ClusterMedoid と提案手法は同じ結果盤面のクラスタを用いており，違いは代表手の選び方，採用するクラスタの数，評価のルールの 3 点である（表 1）．同じ外乱込み評価のルールで比べると，提案手法の regret は既存手法より 0.087（22%）小さかった．既存手法の代表手は価値を見ないため，約 14 個の子を持っても，同じ数を無作為に選ぶ場合と性能が変わらなかった．これに対して提案手法は，クラスタ価値で 4 個のクラスタを選び，その中で推定値が最大の手を代表とすることで，子の数を減らしながら有望な手を残している．差が大きいのは最良手のクラスタが大きい局面であり，大きなクラスタの中から盤面上の中心ではなく価値の高い手を選ぶことが，既存手法との差を生んでいる．これは，先行研究 [7] で課題とした代表点の選び方と大きなクラスタの扱いに対する一つの解決である．また，先行研究では全探索との一致率で評価していたが，$P = 200$ における外乱込み全探索 AllGrid 自体の regret は 0.334 であり，全探索の選択を正解とみなすことはできない．審判値に基づく regret で評価したことで，提案手法が全探索を上回ることを示せた．

5.2 人間にとって意味のあるクラスタ

提案手法のクラスタは，探索の単位であるだけでなく，人間が盤面を見て理解できる単位にもなりうる．**図 4** に相手の中心の石をはじき出す手が最良となる二つの盤面について, それぞれ最良手 A と別の道 B の異なる投げ方の結果盤面を示す．図の上段と下段では, それぞれ異なる初期盤面からの結果盤面が生成されるまでの流れを示し, いずれの場合も赤チームの投げる盤面である．図の構成は左から (a) 初期状態, (b) 全候補の着地点分布, (c) 最良手 A を投げ終わった盤面, (d) 別の道 B を投げ終わった盤面である．図 a の石上の数字はハウスの中心から近い順に割り当てられている．図 b にある色はそれぞれクラスタを区別している．緑の円は最良手 A の目標投球位置を示し, 黒い三角は別の道 B の目標投球位置を示している．図 c にある黒いバツはハウス外へ移動した石のことを表し, 投げた石は白い点がついている．

2 つの局面はどちらも相手（青）が No.1 ストーンを持ち，手番の赤が負けている局面である．最良手 A はいずれも相手のティー付近の石に弱く当てる手であり，投げた直後の盤面で数えた得点（以下，即時得点）を，局面 1 では相手の石 2 個（と奥の自分の石 1 個）をプレーの外に出して −1 から +4 に，局面 2 では相手の No.2 を出して No.1 を奥へ押しやり −2 から +2 にする (図 c)．これらの手の結果盤面は他の候補と似ておらず，最良手のクラスタは局面 1 で 2 手，局面 2 で 1 手だけからなる．

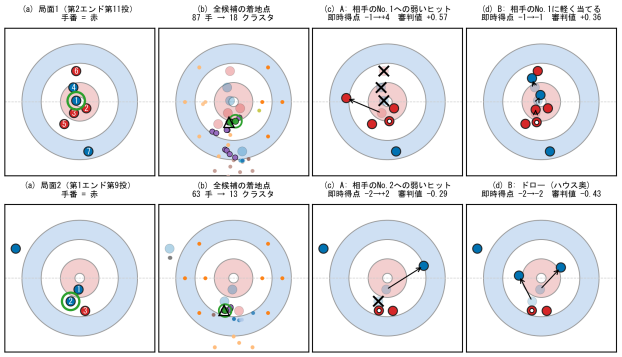


図 4 相手の中心の石を取り除く最良手の例 (上段：局面 1，下段：局面 2

図 b のように，A の投げた石は別のクラスタの手とほぼ同じ位置に止まるが，他の石の動き方が異なるため別のクラスタに分かれている．すなわちクラスタは，狙いや投げた石の止まる位置ではなく投げた後の盤面全体で手をまとめており，相手の石を取り除く手が多数のドローやガードに埋もれずに現れる．

一方，B は即時得点を変えないが，審判値は最良に近い手である（局面 1 で +0.36，局面 2 で −0.43．最良との差は 0.21 と 0.14）．局面 2 では，即時得点で 2 点を得る A でもエンドの終わりまで考慮した期待値は −0.29 にとどまり，即時得点の差ほど価値の差は大きくない．

クラスタ価値の順に結果盤面を並べれば，即時得点だけでは気付きにくい同等に有利な別の手も人間に示せる．

5.3 ボロノイ領域との関係

VOO [5] は，評価済みの行動を母点として行動空間をボロノイ分割し，最良の母点のセルの中を重点的に探索する．提案手法も，候補を領域に分けて各領域を 1 つの代表で扱い，推定値が最大の候補を含むクラスタを必ず採用する．各候補を最も近い代表に割り当てた領域を，結果盤面の距離による一般化したボロノイ領域とみなせば，提案手法は VOO と同様に最良の領域を重点的に探す方法と位置づけられる．ただし距離 d はベクトル空間の距離ではなく，平均連結法のクラスタは厳密にはボロノイ領域ではない．両者の本質的な違いはどの空間で近さを測るかにある．そこで，VOO が前提とする「近い行動は近い価値を持つ」という性質が行動空間と結果盤面の空間のどちらでよく成り立つかを調べた．

図 5(a) の局面 3（第 7 エンド第 10 投，青の手番）では，

© 2026 Information Processing Society of Japan

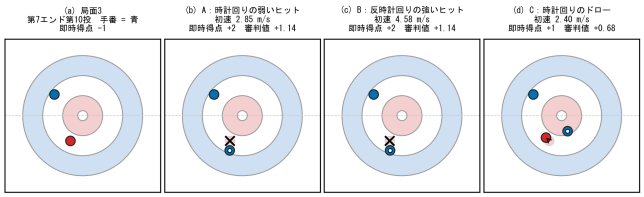


図 5 行動は大きく違うが結果盤面がほぼ同じになる 2 手の例 (局面 3)．(a) 局面．(b) 手 A．(c) 手 B．A と B は回転方向が逆で初速も 1.7 m/s 違うが，どちらも赤の石をプレーの外に出して投げた石がその位置に止まり，同じクラスタに入る．(d) 手 C．行動は A に近いが結果が異なり，別のクラスタに入る．矢印は動いた石，×は外に出た石，白点は投げた石を表す

ティーの左手前 0.85 m に赤の石があって No.1 ストーンとなり，左奥 1.08 m に青の石がある．手 A と手 B はどちらもこの赤の石へのヒットであるが，A は時計回りに初速 2.85 m/s で弱く当て，B は反時計回りに初速 4.58 m/s で強く当てる．回転方向が逆で初速も 1.7 m/s 違うため，行動としては大きく離れている．それにもかかわらず，どちらも赤の石をプレーの外に出し，投げた石がその位置にほぼ止まるため，結果盤面はほぼ同じになり（(b)，(c)），審判値もともに +1.14 である．

一方，手 C は時計回りのドローであり，初速 2.40 m/s と A に近く回転方向も同じなので，行動としては B よりもはるかに A に近い．しかし C は赤の石に軽く触れてティーの近くに止まり，赤の石を残すため（(d)），審判値は +0.68 と A より 0.46 低い．提案手法の距離でも，A と B の結果盤面はこの局面の候補の組の中で最も近い 2%に入り，A と C は中央値より遠い．すなわちこの局面では，価値が近い手は行動が近い手ではなく結果盤面が近い手であり，提案手法は A と B を同じクラスタにまとめている．

以上から，カーリングでは VOO が前提とする「近い手は価値も近い」という性質は，行動の近さではなく結果盤面の近さについて成り立ちやすい．提案手法は，VOO の「最良の領域を重点的に探す」という考え方を，行動空間ではなく結果盤面の空間で用いたものと位置づけられる．クラスタ価値（クラスタ内の推定値の平均）で領域の有望さを比べられるのも，同じクラスタの手の価値がそろっているためである．

6. 今後の課題

実験から，提案手法には次の弱みがあることが分かった．まず，エンドの最終投では事前推定による選別で得点につながる少数の手を落とし，全探索より性能が下がる．最終投の局面だけ全探索に切り替えると，$P = 200$ の 500 局面全体の regret は計算上約 0.02 下がる．次に，プレイアウト回数が多くなると全探索との regret の差は小さくなる．提案手法の利点はプレイアウト回数が少ない場合に限られるため，持ち時間や探索の計算量に応じてクラスタリングを使用するか否かを検討する必要がある．さらに，得点期待値に基づく選別はルートノードだけに適用しており，深さ 1 以降のノードでは既存手法の規則を用いている．事前推定のコストを抑えながら，内部のノードにも価値に基づく選別を広げることが課題である．

7. おわりに

本論文では，デジタルカーリングの MCTS における候補手の削減について，結果盤面のクラスタを保ったまま，クラスタの採用と代表手の選出に得点期待値を用いる ClusterValue と，外乱込み評価を組み合わせた手法を提案した．審判値に基づく regret で 500 局面を同じプレイアウト回数で比較した結果，提案手法は既存手法 ClusterMedoid より regret が 22%小さく，候補を絞らない全探索も上回った．その利得は計算プレイアウト回数が少ないほど大きく，最終投を除けば全探索の約 5〜6 割の計算量で同程度の品質が得られた．性能の向上は，得点期待値に基づく選別と外乱込み評価によるものであり，medoid による代表手の選び方は無作為な選び方と変わらなかった．また，結果盤面のクラスタは手を結果の種類でまとめ，相手の中心の石を取り除くような他と似ていない最良手や，一時的な得点は低いがエンド全体では同等に有利な別の道を，独立したまとまりとして示せることを確認した．今後は最終投での選別の改善と，持ち時間のある対戦および人間による評価を可能にすることを目標とする．

参考文献

[1] 北清勇磨，岡田雷太，伊藤毅志：デジタルカーリングサーバーの提案と紹介，情報処理学会研究報告ゲーム情報学，Vol. 2014, No. 2, pp. 1–5 (2014).
[2] 大渡勝己，田中哲朗：モンテカルロ木探索によるデジタルカーリング AI，第 21 回ゲームプログラミングワークショップ 2016 論文集，pp. 180–187 (2016).
[3] 加藤 修，飯塚博幸，山本雅人：不確定性を含むデジタルカーリングにおけるゲーム木探索，情報処理学会論文誌，Vol. 57, No. 11, pp. 2354–2364 (2016).
[4] Yee, T., Lisý, V. and Bowling, M.: Monte Carlo Tree Search in Continuous Action Spaces with Execution Uncertainty, *Proc. 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)*, pp. 690–696 (2016).
[5] Kim, B., Lee, K., Lim, S., Kaelbling, L. P. and Lozano-Pérez, T.: Monte Carlo Tree Search in Continuous Spaces Using Voronoi Optimistic Optimization with Regret Bounds, *Proc. 34th AAAI Conference on Artificial Intelligence (AAAI-20)*, pp. 9916–9924 (2020).
[6] Kwak, Y., Hwang, I., Kim, D., Lee, S. and Zhang, B.-T.: Efficient Monte Carlo Tree Search via On-the-Fly State-Conditioned Action Abstraction, *Proc. 40th Conference on Uncertainty in Artificial Intelligence (UAI 2024)*, PMLR, Vol. 244, pp. 2076–2093 (2024).
[7] 仲亜斗夢，横山大作：デジタルカーリングにおける MCTS 候補手削減のためのクラスタリング適用の試み，情報処理学会研究報告ゲーム情報学 (GI)， Vol. 2026-GI-58 (2026).

© 2026 Information Processing Society of Japan