Ancestor-Based α-β Bounds for Monte-Carlo Tree Search*

Tom Pepels[1 [0009-0007-1437-0693]] and Mark H.M. Winands[1 [0000-0002-0125-0824]]

Department of Advanced Computing Sciences, Maastricht University,
Maastricht, The Netherlands
{tom.pepels, m.winands}@maastrichtuniversity.nl

1 α-β Bounds for Monte-Carlo Tree Search

We propose an enhancement that addresses a shortcoming of Monte-Carlo Tree Search (MCTS) [1], which traditionally considers each node as a standalone multi-armed bandit (MAB). This isolated approach neglects results from other parts of the search tree, potentially limiting the effective use of accumulated data. Inspired by minimax, this approach integrates ancestral bounds (α and β) into UCT [2], allowing them to intersect and be dynamically adjusted. By incorporating path-dependent information and refining predictions based on evaluations, our approach enhances decision making and improves selection accuracy, thus moving beyond viewing tree nodes as independent MAB challenges.

$$\alpha = \max\left(\alpha, V(s,a) - CB(s,a)\right)$$
$$\beta = \min\left(\beta, V(s,a) + CB(s,a)\right)$$

Before selection, α and β are dynamically adjusted using the value function $V(s,a)$ and confidence bounds $CB(s,a)$ (as in UCT). Whenever these bounds are updated, the corresponding lower bound α_- and the upper bound β_+ are stored for use in subsequent decisions. This leads to a constant revision of the best current upper and lower bounded values for the maximizing player along the chosen path.

Our goal is to enhance UCT in MCTS by using the α and β bounds to better balance exploration and exploitation. These bounds guide UCT in deciding whether to focus on known promising paths or explore new, potentially better options. When $\alpha < \beta$, the algorithm prioritizes exploitation of promising paths, whereas when $\alpha > \beta$, it indicates deviation towards exploring alternative options with potential better results. To achieve this, the UCT formula is modified by incorporating the following:

$$\delta_{\alpha\beta} = 1 - (\beta - \alpha) \cdot (1 - (\beta_+ - \alpha_-)), \tag{1}$$

$$\varDelta_{\alpha\beta} = C_{\alpha\beta}^2 \cdot \ln(\delta_{\alpha\beta} \cdot N(s)), \tag{2}$$

$$\text{argmax}_a \left(\frac{Q(s,a)}{N(s,a)} + C\sqrt{\frac{\varDelta_{\alpha\beta} \cdot \ln N(s)}{N(s,a)}} \right) \tag{3}$$

* Publication: Pepels, Tom, and Winands, Mark H. M. Ancestor-Based α-β Bounds for Monte-Carlo Tree Search. *2024 IEEE Conference on Games (CoG)*, pp. 1-4. [5]

In Formula 2, the parameter $C_{\alpha\beta}$ is introduced to modify the impact of $\Delta_{\alpha\beta}$ on the exploration process, in our experiments not deviating from C by more than 0.2. $\delta_{\alpha\beta}$ (Formula 1) measures the disparity between β and α, reflecting the uncertainty in their limits. As this uncertainty increases, or as β and α converge, $\delta_{\alpha\beta}$ decreases, which can even become negative if $\beta < \alpha$, emulating minimax pruning. The adjustment of the exploration term using $\Delta_{\alpha\beta}$ is further controlled by $N(s)$, assuring a stronger impact on well-explored nodes and requiring substantial evidence before altering the exploration strategy. When $\Delta_{\alpha\beta} = 0$, or when β or α are undefined, the standard UCT selection formula is used. Otherwise, the modified Formula 3 is applied.

2 Experiments

Our experiments tested MCTS$_{\alpha\beta}$ on four board games selected for their distinct challenges. The experimental configuration incorporated improvements to achieve a higher level of play, such as early playouts [4], and informed playouts. Table 2 presents our approach combined with Implicit Minimax [3].

The experimental results indicate that MCTS$_{\alpha\beta}$ significantly improves win rates in Breakthrough and Mini Shogi, where endgame strategies are critical, demonstrating its effectiveness in scenarios with limited paths to victory. In Amazons, the high branching factor likely limits search depth, reducing the impact of α and β initialization. For Gomoku, the benefits of MCTS$_{\alpha\beta}$ are notable only when combined with implicit minimax, suggesting that this combination is essential for navigating complex strategic options. With sufficient computational resources, Gomoku's performance improves significantly, and similar gains might be achievable in Amazons with further resource allocation.

The efficiency of MCTS$_{\alpha\beta}$ is dependent on the strategic depth of the game and the magnitude of the search space. Games such as Breakthrough and Minishogi, which demand tactical play, gain considerable advantages from $\alpha - \beta$ bounds. In contrast, Amazons and Gomoku, with many similar strategies, see fewer benefits. Future work will explore improving performance in these areas and investigate the theoretical convergence of the approach.

Game	Simulation budget	MCTS$_{\alpha\beta,imm}$ vs. MCTS$_{imm}$	MCTS$_{\alpha\beta}$ vs. MCTS
		Win Rate ± 95% c.i.	Mean ± 95% c.i.
Amazons (8 × 8)	75,000	50.99 ± 2.88	52.08 ± 2.80
	225,000	50.71 ± 2.83	**53.34 ± 2.88**
Gomoku	75,000	52.69 ± 2.75	50.88 ± 2.77
	225,000	**53.57 ± 2.79**	51.16 ± 2.77
Breakthrough	75,000	**53.35 ± 3.07**	52.44 ± 3.03
	225,000	**54.30 ± 3.02**	**55.37 ± 3.07**
Mini Shogi	25,000	**57.16 ± 3.05**	**54.61 ± 2.99**
	75,000	**59.94 ± 3.15**	**58.24 ± 2.97**

References

1. Coulom, R.: Efficient selectivity and backup operators in Monte-Carlo tree search. In: van den Herik, H.J., Ciancarini, P., Donkers, H.H.L.M. (eds.) 5th International Conference on Computers and Games. LNCS, vol. 4630, pp. 72–83. Springer-Verlag (2007)
2. Kocsis, L., Szepesvári, C.: Bandit-based Monte Carlo planning. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) 15th European Conference on Machine Learning. LNCS, vol. 4212, pp. 282–293. Springer-Verlag (2006)
3. Lanctot, M., Winands, M.H.M., Pepels, T., Sturtevant, N.R.: Monte carlo tree search with heuristic evaluations using implicit minimax backups. In: 2014 IEEE Conference on Computational Intelligence and Games. pp. 1–8. IEEE (2014)
4. Lorentz, R.: Amazons discover Monte-Carlo. In: Proceedings of the 6th International Conference on Computers and Games (CG). LNCS, vol. 5131, pp. 13–24 (2008)
5. Pepels, T., Winands, M.H.: Ancestor-based α-β bounds for Monte-Carlo Tree Search. In: 2024 IEEE Conference on Games (CoG). pp. 1–4 (2024). https://doi.org/10.1109/CoG60054.2024.10645672