JIC JORNADA DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA XV EDIÇÃO
Integridade Científica e Combate à desinformação
20 a 22 de outubro
uffs.edu.br/jic
UNIVERSIDADE FEDERAL DA FRONTEIRA SUL

MODELAGEM DE OPONENTES PARA O TREINAMENTO DE MODELOS BASEADOS NA BUSCA EM ÁRVORE DE MONTE CARLO NO JOGO DE TABULEIRO CITADELS

ANDREI CARLESSO CAMILOTTO[1][2], DJONATAN RIQUELME CLEIN BONELLI[3], EDUARDO VINICIUS PERISSINOTTO FIORENTIN[4], JOAO LUIS ALMEIDA SANTOS[5], FELIPE GRANDO[6]

1 Introdução

A Busca em Árvore de Monte Carlo (Monte Carlo Tree Search – MCTS) consolidou-se como uma técnica eficaz para tomada de decisão em jogos de tabuleiro, combinando adaptabilidade e baixo uso de conhecimento específico de domínio. Seu desempenho é notável em jogos determinísticos com informação perfeita, como Go (SILVER et al., 2016). Entretanto, sua aplicação encontra limitações significativas em ambientes com múltiplos jogadores, informação oculta e restrições de tempo, nos quais o alto custo computacional das simulações e a dificuldade de adaptação comprometem a responsividade e a qualidade das decisões (POWLEY et al., 2014).

Neste contexto, abordagens híbridas que integram MCTS e Aprendizado por Reforço (Reinforcement Learning – RL) têm se mostrado promissoras, ao permitir que a fase de planejamento seja desacoplada da execução. A construção de estruturas de decisão persistentes, com armazenamento prévio da busca, viabiliza a seleção instantânea de ações durante o jogo, reduzindo a necessidade de simulações online e preservando a profundidade estratégica (BROWNE, 2012; SWIECHOWSKI, 2023).

Nessa pesquisa foi testado e desenvolvido um modelo híbrido de MCTS com técnicas *offline* de aprendizado por reforço. O modelo foi treinado e testado usando diferentes oponentes em um ambiente de simulação do jogo de tabuleiro Citadels (FAIDUTTI, 2016).

2 Objetivo Geral

O objetivo geral desta pesquisa é desenvolver um framework híbrido MCTS–RL capaz de operar de forma eficiente em jogos de tabuleiro multiplayer com informação oculta,

[1] Acadêmico do curso de Ciência da Computação, UFFS, *campus* Chapecó, andrei.camilotto@gmail.com
[2] Grupo de Pesquisa: Inovação e Desenvolvimento Tecnológico - IDT
[3] Acadêmico do curso de Ciência da Computação, UFFS, *campus* Chapecó
[4] Acadêmico do curso de Ciência da Computação, UFFS, *campus* Chapecó
[5] Acadêmico do curso de Ciência da Computação, UFFS, *campus* Chapecó
[6] Doutor em Computação, UFFS, *campus* Chapecó, **Orientador**.

utilizando Citadels como estudo de caso. A proposta busca eliminar gargalos computacionais do MCTS tradicional por meio da construção offline de uma estrutura de decisão persistente, permitindo ações rápidas e robustas em tempo de execução.

3 Metodologia

Essa pesquisa caracteriza-se por ser de natureza experimental aplicada com análise quali-quantitativa.

A primeira etapa envolveu o desenvolvimento e adaptação de um ambiente de simulação digital do jogo de tabuleiro Citadels. Seguiu-se com o desenvolvimento de 5 agentes especialistas com estratégias customizadas e desenvolvidas para imitarem estratégias reais usadas por jogadores humanos no jogo e o desenvolvimento de um agente de base que apenas escolhe as ações aleatoriamente. Depois de ser feita a validação do ambiente e dos agentes desenvolvidos iniciou-se o desenvolvimento do framework híbrido MCTS-RL.

O modelo Monte Carlo Tree Search (MCTS) é, em sua essência, um algoritmo de busca heurística que opera através da execução de múltiplas simulações. Para cada estado de jogo distinto visitado durante essas simulações, um nó correspondente é criado em uma estrutura de dados em árvore. O estado inicial do jogo representa o nó raiz (root) dessa árvore.

Contudo, a proposta de armazenar toda a árvore e preservar para decisões futuras é inviável por conta do gigantesco fator de ramificação, portanto, para tornar isso possível será necessário abstrair a árvore. Para isso o estado do jogo é dividido em suas características fundamentais (e.g., ouro do jogador, número de cartas na mão, etc.), e cada característica é mapeada para uma tabela de dados independente.

Nessa estrutura, as linhas de uma tabela representam os possíveis valores de uma característica de estado (e.g., os possíveis valores de ouro que um jogador pode ter). As colunas, por sua vez, representam as ações possíveis que o agente pode tomar (e.g., comprar ouro, comprar cartas, entre outras). Na tabela são armazenados dois valores: o número de vitórias alcançadas (n) após tomar aquela ação naquele estado, e o número total de vezes que essa combinação foi explorada (m). Essas tabelas, quando preenchidas através de simulações de partidas do jogo Citadels, configuram o modelo aprendido do agente.

O processo de treinamento também inclui um componente de aprendizado por reforço, após cada partida simulada, os resultados atualizam as tabelas de decisão, influenciando escolhas futuras. Vitórias atuam como recompensas e reforçam ações eficazes, aumentando sua probabilidade ser escolhida novamente, enquanto derrotas reduzem essa possibilidade. Assim,

ao longo de muitas simulações, o modelo adapta-se continuamente, equilibrando exploração de novas alternativas e aproveitamento de estratégias já comprovadas, convergindo gradualmente para comportamentos mais eficazes.

4 Resultados e Discussão

Para avaliar a performance do modelo MCTS proposto, foi crucial entender a dinâmica de poder entre as estratégias base (agentes especialistas e agente aleatório). Para isso, cada uma foi submetida a uma rodada de 10.000 partidas contra quatro oponentes idênticos. O mapa de calor apresentado na Figura 1 ilustra o desempenho de cada estratégia (eixo Y) contra os diferentes adversários (eixo X). Fica evidente que as estratégias possuem forças e fraquezas distintas: os Especialistas 1, 2 e 3 mostram-se competentes na maioria dos cenários, enquanto os Especialistas 4 e 5 são consideravelmente menos eficazes. Como esperado, a estratégia totalmente aleatória apresenta um desempenho muito baixo contra qualquer oponente minimamente estruturado.

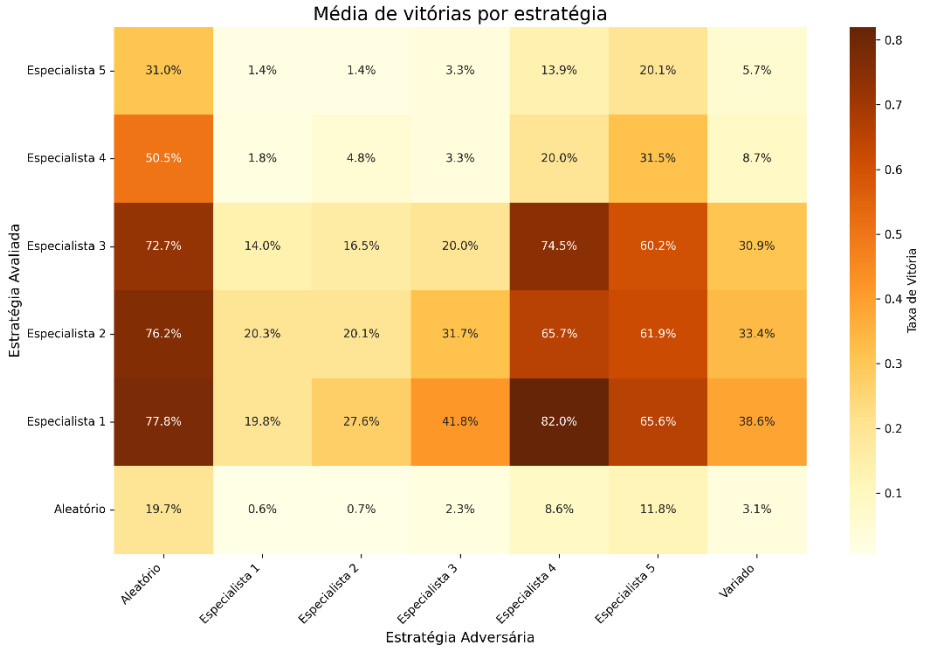

Figura 1: Comparativo entre estratégias

O passo seguinte foi avaliar o modelo MCTS-RL. Realizamos diferentes sessões de treinamento, cada uma focada em um dos oponentes, totalizando 100 mil partidas de treino para cada versão do modelo. Em seguida, o desempenho de cada modelo treinado foi medido contra todos os tipos de adversários seguindo o mesmo modelo de teste anterior, e contra oponentes variados sorteados aleatóriamente. O mapa de calor mostrado na Figura 2 apresenta a taxa de vitória consolidada de cada modelo.

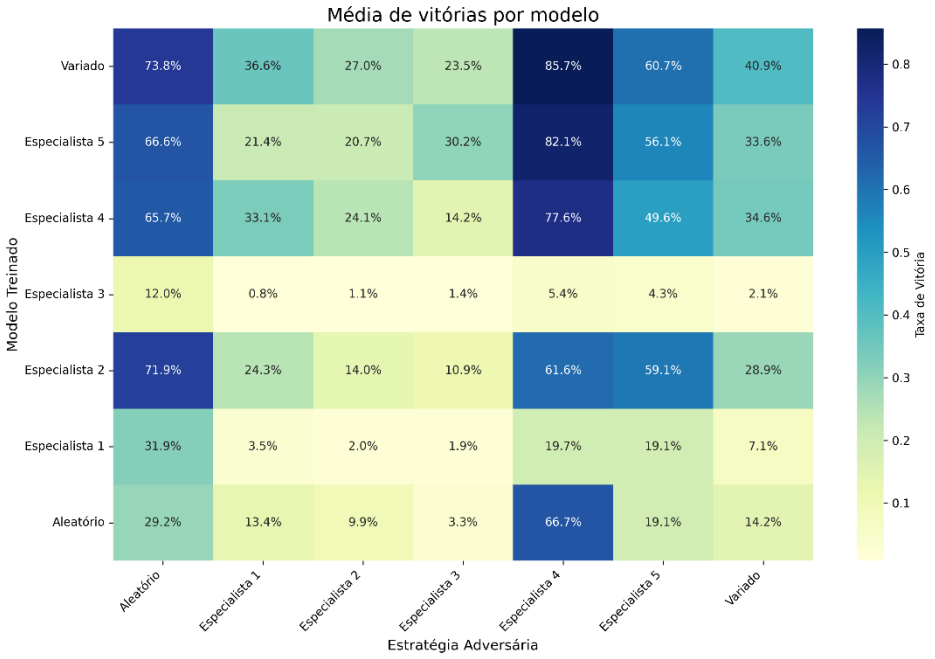

Figura 2: Comparativo entre os modelos treinados

O resultado mais expressivo foi o do modelo treinado no ambiente variado. Este agente obteve um desempenho consistente, superando as outras versões do modelo em quase todos os cenários de teste. Isso sugere que a exposição a uma diversidade de estratégias durante o treinamento foi fundamental para o desenvolvimento de uma capacidade de jogo mais robusta e generalista. Desse modo, ele aprendeu princípios mais amplos de Citadels, tornando-se um jogador mais completo.

Os modelos treinados contra os Especialistas 4 e 5, considerados os adversários mais "fracos", tornaram-se agentes surpreendentemente eficazes. Por outro lado, o treinamento contra os Especialistas 1 e 3, oponentes tidos como "fortes", produziu modelos com um desempenho geral muito pobre. Esse fenômeno indica como o tipo de ambiente de treinamento afetou a exploração de estados pelo modelo. Uma possível explicação para esse contraste está no fato de que estratégias especialistas avançadas tendem a ser altamente especializadas, enquanto o modelo MCTS-RL, no início do treinamento, se aproxima do comportamento Aleatório. Contra oponentes fortes, mesmo ações potencialmente vantajosas podem não resultar em vitórias suficientes para que o modelo reconheça padrões de sucesso. Já diante de adversários menos especializados, as vitórias ocorrem com maior frequência, o que favorece o aprendizado. Ainda assim, trata-se apenas de uma hipótese que necessita de confirmação.

5 Conclusão

Este estudo demonstrou a viabilidade da abordagem proposta, que utiliza uma abstração da árvore do MCTS para preservar o conhecimento adquirido entre as jogadas. Os resultados confirmaram que o modelo é capaz de desenvolver estratégias robustas para o ambiente. O desempenho superior do agente treinado no "Ambiente Variado" evidencia que a exposição a múltiplas táticas é fundamental para a construção de um modelo generalista. Adicionalmente, a pesquisa revelou a conclusão de que a qualidade do agente final depende mais da diversidade do que da força aparente dos oponentes de treinamento. Conclui-se, portanto, que a estrutura de MCTS com memória persistente é uma alternativa promissora, cuja eficiência é maximizada por um treinamento em cenários heterogêneos.

Para possíveis estudos futuros, seria interessante explorar com mais precisão os motívos reais das estratégias menos especializadas resultarem em modelos mais eficientes do que estratégias mais fortes. Além disso, uma comparação do modelo proposto com outros modelos de MCTS também ajudaria a ter uma avaliação mais precisa nossa abordagem.

Referências Bibliográficas

BROWNE, C. B. et al. **A survey of monte carlo tree search methods**. IEEE Transactions on Computational Intelligence and AI in Games, v. 4, n. 1, p. 1–43, 2012.

FAIDUTTI, B. **Citadels**. Deluxe ed. Roseville, MN, USA, 2016. Rulebook.

POWLEY, E. J.; COWLING, P. I.; WHITEHOUSE, D. **Information capture and reuse strategies in monte carlo tree search, with applications to games of hidden information. Artificial Intelligence**, v. 217, p. 92–116, 2014. ISSN 0004-3702.

SILVER, D. et al. **Mastering the game of go with deep neural networks and tree search**. Nature, v. 529, n. 7587, p. 484–489, Jan 2016. ISSN 1476-4687.

SWIECHOWSKI, M. et al. **Monte carlo tree search: a review of recent modifications and applications**. Artificial Intelligence Review, v. 56, n. 3, p. 2497–2562, 2023. ISSN 1573-7462.

Palavras-chave: Busca em árvore de Monte Carlo; Aprendizado por reforço; Agentes; Jogos de tabuleiro.

Nº de Registro no sistema Prisma: PES 2024-0256

Financiamento

UNIVERSIDADE FEDERAL DA FRONTEIRA SUL