Inteligência Artificial de Poker Pluribus

Pedro Tavares de Carvalho

30 de janeiro de 2022

Resumo

Nesse projeto, iremos discutir a inteligência artificial de poker chamada *Pluribus* [4], seus detalhes de implementação e design, além de seus impactos na sociedade e no futuro de IA (Inteligência Artificial) em jogos. Abordaremos o poker como um jogo em teoria de jogos, seu processo de modelagem e os desafios existentes ao se tratar de um jogo como o poker em uma inteligência artificial.

Introdução O poker é um dos jogos mais antigos e mais jogados do mundo. Com regras simples de se entender, porém difíceis de se dominar, o poker atende jogadores casuais e aficionados da mesma forma.

Apesar disso, a revolução das inteligências artificiais, que inclui damas [12], xadrez [6] e mais recentemente GO [13], atingiu diversos outros jogos antes do mesmo. Muito disso é por causa da dificuldade de se modelar uma estratégia ótima em poker.

A história da inteligência artifical no poker é longa, incluindo diversas estratégias, dentre elas algoritmos genéticos [7] e aprendizado profundo [10]. Uma versão mais simplificada do poker, onde os jogadores possuem limites em quanto eles podem apostar, já foi fechada matematicamente [2].

Iremos explorar no que o Pluribus difere das estratégias empregadas até o momento, e o motivo dessas estratégias serem menos eficientes e menos efetivas do que as novas, modelando o poker matematicamente e observando os detalhes que fizeram o Pluribus quebrar a barreira do poker sem limites com 6 jogadores.

Modelagem do Poker em Teoria de Jogos

O poker pode ser modelado em teoria de jogos como um jogo:

De informação incompleta Os agentes do jogo não possuem acesso a quais cartas os seus adversários estão na mão, ou seja, com as informações que ele tem no momento, não é possível determinar com certeza em qual estado de jogo o mesmo está.

Adversarial Os agentes jogam com o objetivo de ganhar, sem colaborar entre si.

Soma zero Quando um dos agentes joga, uma combinação dos outros perde, tornando o valor total existente no jogo fixo.

Estocástico Ações dentro do poker podem produzir estados diferentes, e, consequentemente, valores diferentes.

Em geral, o poker difere de jogos como xadrez e GO por alguns motivos, incluindo o fato de ele ser um jogo de informação incompleta e estocástico, e pelo fato de o mesmo ser jogado com mais de dois jogadores, o que torna a estratégia e a busca dentro da árvore de estados mais complexa do que um minimax simples.

Equilíbrio de Nash e Poker GTO O poker, quando jogado por pessoas reais, é dividido em duas vertentes principais, a vertente exploitativa, que tenta se aproveitar de erros na estratégia do seu adversário, e o poker Ótimo em Teoria dos Jogos (GTO), que tenta maximizar matemáticamente a estratégia, tentando atingir um Equilíbrio de Nash [11].

Em poker sem limites, não existe uma estratégia de Nash fechada, o que é feito são tentativas de aproximação dessa estratégia, no caso do poker humano, com conceitos como valor esperado, Independant Chip Modelling [8] e outros conceitos que ajudam a aproximar a melhor ação em cada momento. Já em inteligências artificiais, a técnica mais usada é a de CRM (Minimização de Arrependimento Contrafactual), que atualiza a árvore de decisões tentando diminuir a quantidade de decisões erradas.

O Pluribus por alto

O Pluribus é composto de três peças principais, abstração de informação, o croquê de estratégia construído por minimização de arrependimento, e o algoritmo de busca em tempo real que cria estratégia em nós não antes explorados.

A estratégia do Pluribus foi construída, principalmente, por jogos com versões de si mesmo. Essa estratégia é muito utilizada em inteligências artificiais, incluindo o xadrez [14] e GO [13]. Com esses jogos, o Pluribus desenvolve uma estratégia base, que é utilizada e atualizada no momento da busca em tempo real, quando o agente está realmente jogando.

Abstrações de Informação

Para simplificar a busca e melhorar a estratégia da IA, o Pluribus aplica técnicas de simplificação de informação, tanto a nível de apostas possíveis quanto a nível de valor de mão.

Abstração de Apostas A simplificação de apostas age tanto no espaço de apostas con-

sideradas pelo agente em sua jogada quanto em apostas de adversários. O processo é uma discretização das apostas, transformando valores específicos de apostas em um conjunto específico de apostas possíveis.

Abstração de Mãos Esse processo diminui o espaço de mãos possíveis dentro de um jogo de poker. A idéia é que mão de valores similares, como um *flush* com carta alta em 10 e um *flush* com carta alta em 9, serão agrupadas em um mesmo conjunto, e ações similares (se não idênticas) serão tomadas em situações que surgem com essas mãos.

Isso não é uma regra, e dependendo da especificidade da situação, a abstração pode ser desconsiderada, e as mãos podem ser tratadas como diferentes pela estratégia.

Minização de arrependimento contrafactual de Monte Carlo

Para a construção da estratégia base, é utilizada uma técnica chamada de MCCRM (Minimização de Arrependimento Contrafactual de Monte Carlo) [9, 3].

Essa técnica consiste na simulação de jogos da IA contra si mesma, e da revisão das decisões tomadas no jogo completo a fim de minimizar o "arrependimento"do agente, ou seja, aumentando a probabilidade de decisões boas e diminuindo a de decisões ruins.

Diferente da CRM normal, em que a cada geração da estratégia a árvore completa de decisões é atualizada, o MCCRM simula somente um jogo[1] de todos os possíveis, e atualiza as decisões que são geradas nesse jogo somente.

Busca em jogos de informação incompleta

Diferente de jogos de infomação completa, como xadrez e GO, em que você consegue ter certeza do estado em que você está e da estratégia do seu adversário, busca em ambientes de informação incompleta não conseguem se basear apenas na avaliação de um estado na estratégia de Nash, ao menos não de um estado não-folha.

[1] Um jogo é uma descida completa na árvore de decisões, até se chegar em um nó folha.

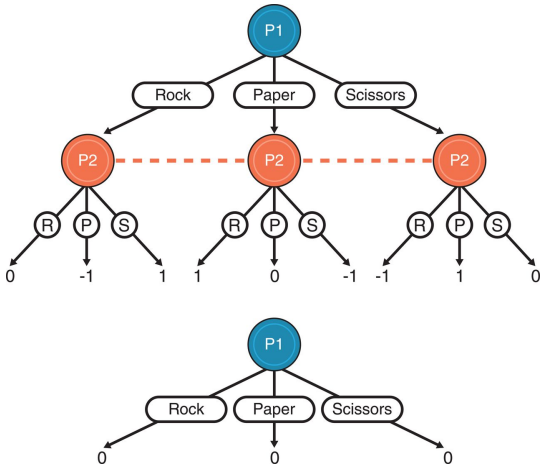

Figura 1: Avaliação e busca em um jogo de pedra papel ou tesoura. Primeiro se vê a descida das decisões e valores possíveis na árvore, depois se vê a avaliação de cada decisão em uma estratégia de Nash.

Isso acontece pois a avaliação assume que o adversário não usará uma estratégia diferente no futuro, ou seja, continuará utilizando uma estratégia de Nash ótima, e portanto tomará as mesmas decisões. Isso pode ser visto na figura 1, onde uma busca em um jogo de pedra papel ou tesoura causa uma estratégia que não implica em otimalidade.

Por exemplo, se tivéssemos um algoritmo guloso, uma busca poderia sempre resultar em *pedra* como ação, o que faria com que se um adversário racional mudasse sua estratégia, ele pudesse explorar os nossos vieses.

Para contornar esse problema, o Pluribus aplica uma técnica de pesquisa com profundidade limitada [5] em que diferentes estratégias adversárias são emuladas e avaliadas, observando o valor em todas elas. Para o Plúribus, foram utilizadas 4 estratégias diferentes, sendo essas vieses construídos a partir da *blueprint* de estratégia gerada com o MCCRM.

Glossário

arrependimento O quanto um agente preferiria ter tomado uma decisão à decisão real dele. i, ii

estratégia de Nash Uma estratégia ótima em um equillíbrio de Nash. Desvios de uma estratégia de Nash, não podem melhorar a performance de um agente. ii

flush Um jogo em poker em que um jogador possui cinco cartas do mesmo naipe. ii

Siglas

CRM Minimização de Arrependimento Contrafactual i

IA Inteligência Artificial i, ii

MCCRM Minimização de Arrependimento Contrafactual de Monte Carlo ii

Referências

[1] *Tuomas Sandholm: Poker and Game Theory*. https://www.youtube.com/watch?v=b7bStIQovcY.

[2] Bowling, M., N. Burch, M. Johanson e O. Tammelin: *Heads-up limit hold'em poker is solved*. Science, 347(6218):145–149, 2015.

[3] Brown, N. e T. Sandholm: *Solving imperfect-information games via discounted regret minimization*. Em *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 1829–1836, 2019.

[4] Brown, N. e T. Sandholm: *Superhuman AI for multiplayer poker*. Science, 365(6456):885–890, 2019.

[5] Brown, N., T. Sandholm e B. Amos: *Depth-limited solving for imperfect-information games*. arXiv preprint arXiv:1805.08195, 2018.

[6] Campbell, M., A. Hoane e F. hsiung Hsu: *Deep Blue*. Artificial Intelligence, 134(1):57–83, 2002, ISSN 0004-3702. https://www.sciencedirect.com/science/article/pii/S0004370201001291.

[7] Carter, R. G. e J. Levine: *An Investigation into Tournament Poker Strategy using Evolutionary Algorithms*. Em *2007 IEEE Symposium on Computational Intelligence and Games*, pp. 117–124, 2007.

[8] Gilbert, G. T.: *The independent chip model and risk aversion*. arXiv preprint arXiv:0911.3100, 2009.

[9] Lanctot, M., K. Waugh, M. Zinkevich e M. Bowling: *Monte Carlo Sampling for Regret Minimization in Extensive Games*. Em Bengio, Y., D. Schuurmans, J. Lafferty, C. Williams e A. Culotta (eds.): *Advances in Neural Information Processing Systems*, vol. 22. Curran Associates, Inc., 2009. https://proceedings.neurips.cc/paper/2009/file/00411460f7c92d2124a67ea0f4cb5f85-Paper.pdf.

[10] Moravcík, M., M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson e M. H. Bowling: *DeepStack: Expert-Level Artificial Intelligence in No-Limit Poker*. CoRR, abs/1701.01724, 2017. http://arxiv.org/abs/1701.01724.

[11] Nash, J. F. *et al.*: *Equilibrium points in n-person games*. Proceedings of the national academy of sciences, 36(1):48–49, 1950.

[12] Samuel, A. L.: *Some Studies in Machine Learning Using the Game of Checkers*. IBM Journal of Research and Development, 3(3):210–229, 1959.

[13] Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel e D. Hassabis: *Mastering the game of Go with deep neural networks and tree search*. Nature, 529(7587):484–489, Jan 2016, ISSN 1476-4687. https://doi.org/10.1038/nature16961.

[14] Silver, D., T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel,

T. P. Lillicrap, K. Simonyan e D. Hassabis: *Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm*. CoRR, abs/1712.01815, 2017. http://arxiv.org/abs/1712.01815.