MỞ RỘNG ALPHA ZERO VÀO CÁC TRÒ CHƠI CÓ THÔNG TIN KHÔNG HOÀN HẢO

PHẠM THĂNG LONG - 240101016

Tóm tắt

- Lớp: CS2205.CH183
- Link Github: https://github.com/Skizdukion/CS2205.CH183
- Link YouTube video:
- Phạm Thăng Long - 240101016

Giới thiệu

Sự thành công của Alpha Zero trong việc tạo ra một thuật toán đã đánh bại con người trong các môn cờ cổ điển như Chess, Go, hay shogi đã đánh dấu một bước ngoặt mới trong lĩnh vực trí tuệ nhân tạo và đặc biệt trong lĩnh vực học tăng cường nói chung.

Tuy nhiên phương pháp này gặp phải thách thức lớn khi áp dụng cho các trò chơi có thông tin không hoàn hảo như Poker hay Liar Dice vì Alpha Zero không có khả năng suy luận thông tin ẩn của đối thủ.

Đề xuất một kiến trúc mới giúp thuật toán có thể hoạt động hiệu quả trên dạng môi trường trên

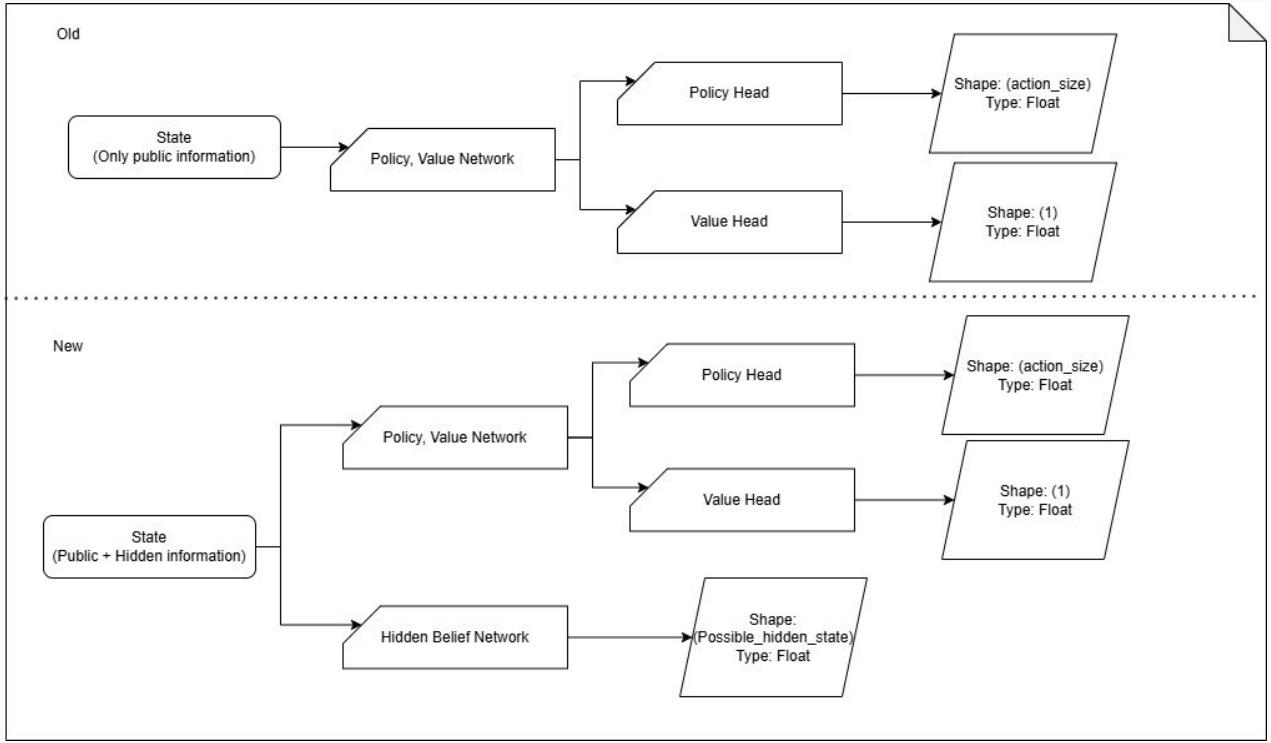
Mục tiêu

- Nghiên cứu, đề xuất phương pháp để cho kiến trúc Alpha Zero có khả năng suy luận thông tin ẩn
- Hiện thực kiến trúc mới và huấn luyện mô hình trên Leduc Poker (một biến thể đơn giản của Poker)
- So sánh đánh giá được kiến trúc mới so với các mô hình hiện tại

Nội dung và Phương pháp

Nội dung 1: Thiết kế và tích hợp kiến trúc mạng nơ-ron có khả năng lý luận thông tin ẩn

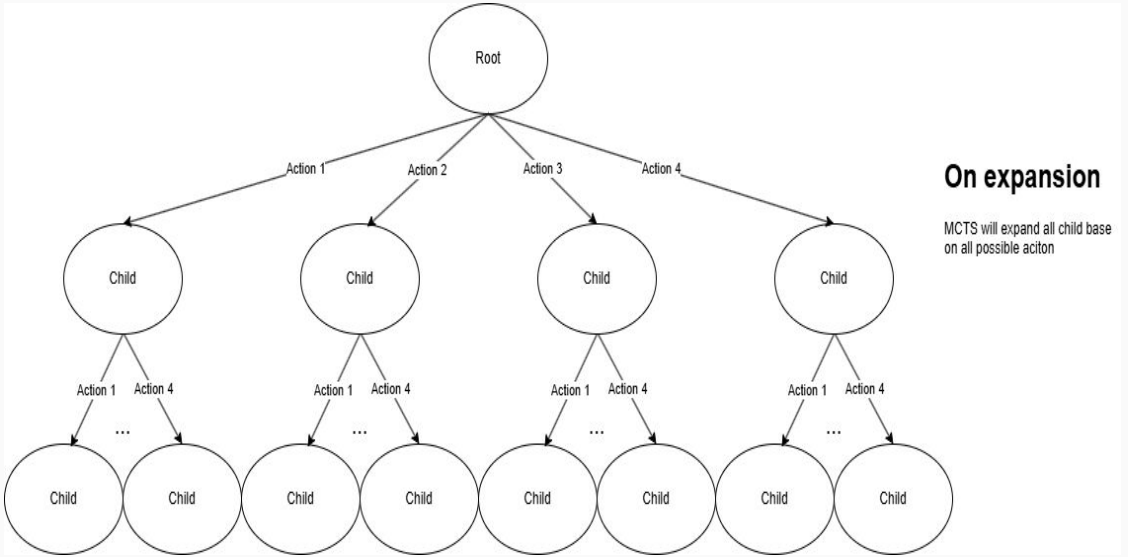
Phương pháp: Đề xuất bằng một kiến trúc mạng neural mới



Nội dung và Phương pháp

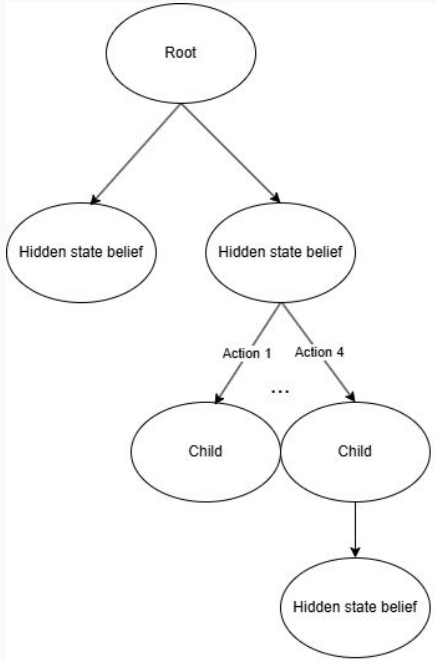
Nội dung 2: Thiết kế, chỉnh sửa thuật toán tìm kiếm cây cho phù hợp với lý luận thông tin ẩn

Phương pháp: Thuật toán cây tìm kiếm mới sẽ được thiết kế để trực tiếp sử dụng thông tin suy luận từ mạng suy luận thông tin ẩn



Thuật toán MCTS trong Alpha Zero gốc

Nội dung và Phương pháp



assume state from belief

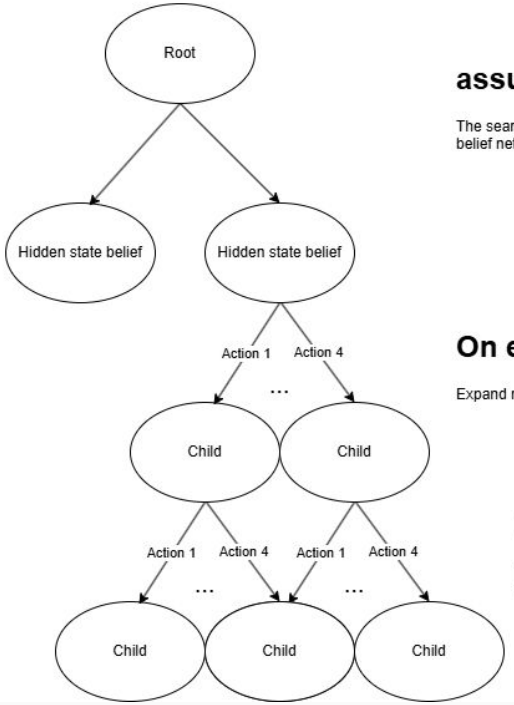
The search will perform sampling from the percentage of hidden belief network to assume the infomation that the opponent have.

On expansion

Expand new child base on all possible action from root node

Extra asume sample

If on current tree there are still hidden information that need to assume, the tree will continue to create new hidden state belief



assume state from belief

The search will perform sampling from the percentage of hidden belief network to assume the infomation that the opponent have.

On expansion

Expand new child base on all possible action from root node

Expand from action

If on current tree there are no more hidden information need to be asume, the tree will expand normally like alpha zero

Thuật toán MCTS mới phù hợp với suy luận thông tin ẩn

Nội dung và Phương pháp

Nội dung 3: Huấn luyện mô hình

Phương pháp:

- Tích hợp cơ chế self play
- Thiết kế hàm loss
- Thực hiện hyperparameters searching

Nội dung 4: Đánh giá kiến trúc mới

Phương pháp:

- Triển khai và thực hiện training một số thuật toán khác
- Thiết kế môi trường cho việc so sánh
- Lưu trữ so sánh chỉ số

Kết quả dự kiến

- Xây dựng được môi trường trò chơi Leduc Poker cho phép người thật tương tác với mô hình
- Kiến trúc đề xuất có thể huấn luyện được mô hình đạt tới Nash Equilibrium trong cơ chế tự chơi ở Leduc Poker
- Mở ra một hướng nghiên cứu mới trong việc huấn luyện mô hình trong môi trường thông tin không hoàn hảo

Tài liệu tham khảo

[1] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. arXiv preprint arXiv:1712.01815.

[2] Zinkevich, M., Johanson, M., Bowling, M., Piccione, C.: Regret minimization in games with incomplete information. In: Advances in Neural Information Processing Systems, pp. 1729– 1736 (2008)

[3] Brown, N., Lerer, A., Gross, S., Sandholm, T.: Deep counterfactual regret minimization. In: International Conference on Machine Learning, pp. 793–802 (2019)

[4] Noam Brown, Anton Bakhtin, Adam Lerer, and Qucheng Gong. Combining deep reinforcement learning and search for imperfect-information games. In Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

[5] H. Fu, W. Liu, S. Wu, Y. Wang, T. Yang, K. Li, J. Xing, B. Li, B. Ma, Q. Fu et al., "Actor-critic policy optimization in a large-scale imperfect-information game," in International Conference on Learning Representations, 2021, pp. 1–12.

[6] Brown N, Sandholm T. Superhuman ai for multiplayer poker. Science. 2019;365:eaay2400.

[7] Lanctot, M., Waugh, K., Zinkevich, M., Bowling, M.: Monte Carlo sampling for regret minimization in extensive games. In: Advances in Neural Information Processing Systems, pp. 1078–1086 (2009)

[8] Moravcˇ´ık, M. et al. Deepstack: Expert-level artificial intelligence in heads-up nolimit poker. Science 356, 508–513 (2017).

[9] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347, 2017.