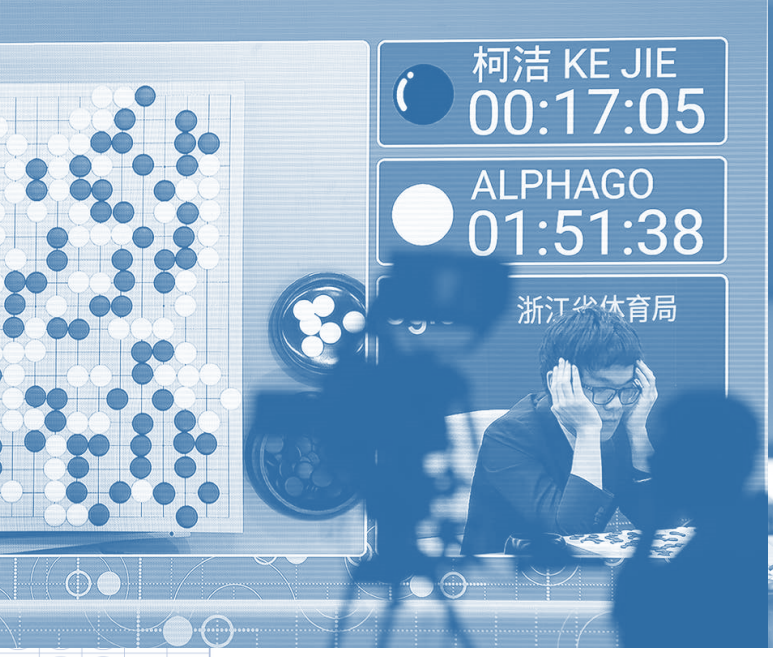
AlphaGo đến AlphaZero: Kẻ hạ bệ trí tuệ con người?

Phần 1: Màn chào sân ấn tượng của Kì thủ vô tri!

Nếu bạn là một người quan tâm đến trí tuệ nhân tạo, hay đam mê với các thể loại cờ, đặc biệt là cờ vây, hẳn bạn đã nghe tới **AlphaGo**, một phần mềm "kì thủ cờ vây" được phát triển bởi công ty nghiên cứu trí tuệ nhân tạo DeepMind.

Năm 2016, AlphaGo đã đánh bại hai kì thủ sừng sỏ thế giới là *Lee Sedol – nhà vô địch cờ vây thế giới 18 lần* với tỉ số 4-1, và chiến thắng toàn diện trước *Ke Jie – nhà đương kim vô địch thế giới cờ vây* với tỉ số 3-0 vào năm 2017.

Vào thời điểm đó, sự kiện gây chấn động giới cờ vây này đã nhanh chóng tràn ngập trang nhất của các đầu báo nổi tiếng nhất, bởi đây không chỉ hạ bệ danh dự của các kì thủ thế giới, mà còn đánh dấu việc trí tuệ máy tính đã vượt xa trí tuệ loài người. Một bộ phim tài liệu cùng tựa đề **AlphaGo** cũng được ra mắt, thuật lại toàn bộ quá trình phát triển AlphaGo và những trận đấu với Lee Sedol. Bộ phim đã đoạt được nhiều đề cử và giải



thưởng tại các liên hoan phim thế giới như Liên hoan phim quốc tế Denver, Liên hoan phim Philadelphia...

Thế nhưng, những người hiểu biết về trí tuệ nhân tạo cũng chỉ tặc lưỡi, cho rằng **AlphaGo** chỉ đơn thuần được học hỏi các nước đi của những kì thủ cờ vây chuyên nghiệp, chứ **không hề trải qua quá trình tự suy nghĩ, tự học nước đi như con người.**

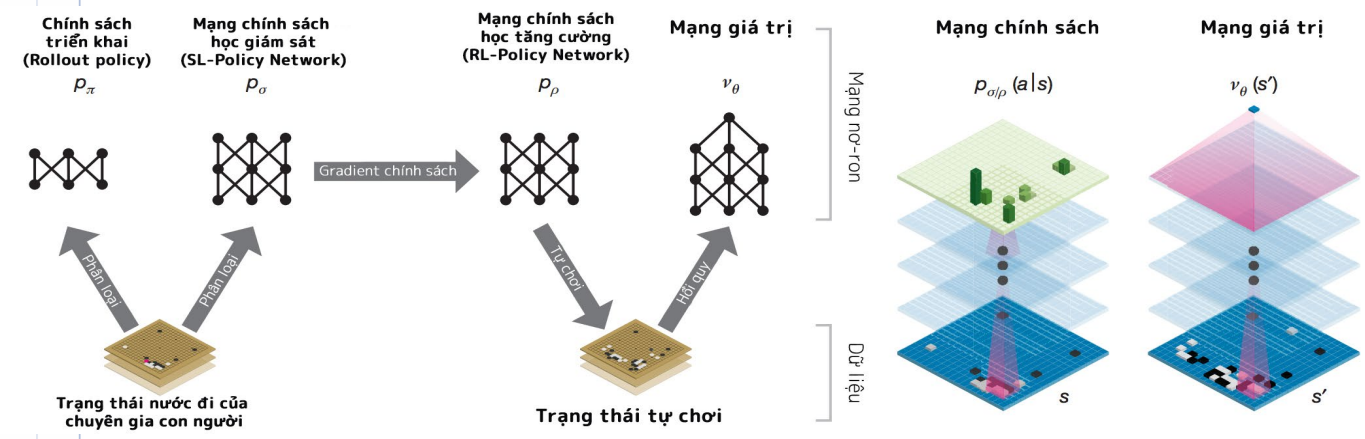
AlphaGo

Quả thật vậy, AlphaGo đã sử dụng 4 Deep Convolutional Neural Network (Mạng nơ-ron tích chập sâu), 3 Policy Network (Mạng chính sách) và 1 Value Network (Mạng giá trị). Trong đó:

- **Supervised Learning Policy Network** (Mạng chính sách học có giám sát): 2 mạng chính sách được học các nước đi "thần thánh" của các kì thủ, hay còn gọi là Imitation Learning (học bắt chước).
- **Reinforcement Learning Policy Network** (mạng chính sách học tăng cường): Mạng chính sách thứ ba này được học tăng cường dựa trên cơ chế self-play (tự chơi). Mạng hiện tại luôn được chơi với một mạng được chọn ngẫu nhiên từ một vài lần lặp trước đó.
- **Rollout policy** (chính sách triển khai) là một mạng nơ-ron nhỏ hơn. Do đó, độ chính xác mô hình hóa các nước đi của nó thấp hơn đáng kể so với mạng dung

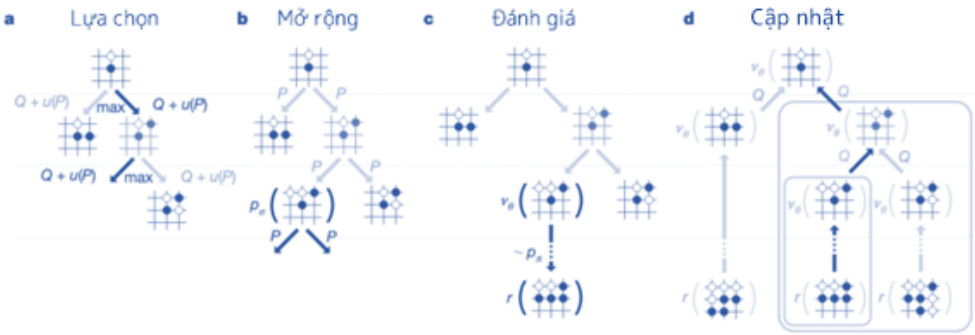
lượng lớn hơn. Tuy nhiên, thời gian suy luận của mạng chính sách triển khai rất ngắn, điều này rất hữu ích cho việc mô phỏng trên cây tìm kiếm Monte Carlo.

- Sau đó, bộ dữ liệu self-play (tự chơi) đào tạo Mạng giá trị để dự đoán người thắng cuộc tại trạng thái hiện tại của trò chơi.



Hình 1. *Các mạng nơ-ron bên trong AlphaGo.*

Cuối cùng, toàn bộ các Mạng trạng thái và Mạng giá trị ở trên được gom lại và đưa vào **Monte Carlo Tree Search** (Cây tìm kiếm Monte Carlo). Cây Monte Carlo hoạt động dựa trên 4 cơ chế chính: Selection (Lựa chọn) – Expansion (Mở rộng) – Evaluation (Đánh giá) – Backup (Cập nhật). Chuỗi nước đi đại diện như một nhánh cây. Nhánh được truy cập nhiều nhất được đánh dấu là nước đi tốt nhất.



Hình 2. *Các bước hoạt động trong cây tìm kiếm Monte Carlo*

Quay lại với trận đấu của AlphaGo với Lee Sedol, ở trận đấu thứ 4, bằng nước đi thứ 78 *"tesuji thần thánh"*, Lee Sedol đã đánh bại AlphaGo. Trận thua của AlphaGo đã trở thành đề tài bàn tán nóng hổi trên khắp các diễn đàn. Họ vạch trần black-box (hộp đen) của AlphaGo **chỉ như một đứa trẻ dễ nổi nóng**.

Quả thật, sau nước đi thứ 78, AlphaGo vẫn tự tin vào tỉ lệ thắng của mình, nhanh chóng chọn ra nước đi tiếp theo có tỉ lệ thắng cao nhất, mà không hề nhận ra mình đã bị Lee Sedol cài vào bẫy. Sau 10 nước, AlphaGo mới nhận ra mình đã bị lừa. Từ nước thứ 87, tỉ lệ thắng được tính toán bởi AlphaGo giảm mạnh, thời gian tính toán các nước đi cũng tăng đột ngột.

Nguyên nhân chính có thể là việc AlphaGo **học vẹt nước đi của con người** quá nhiều, dẫn đến **tự tin thái quá vào khả năng dự đoán** của mình. Một nguyên nhân khác có thể là trong quá trình tìm kiếm của cây Monte Carlo, AlphaGo đã cố gắng cắt bỏ những nhánh trình tự ít liên quan hơn nên với những nước cờ đặc biệt, **trạng thái của ván cờ có thể nằm "ngoài tầm nhắm".**

Và cuối cùng, AlphaGo đầu hàng ở nước 105, đánh dấu chiến thắng cho con người trước trí tuệ nhân tạo. Song cuối cùng thì, AlphaGo vẫn chỉ là một **bộ máy được nhồi nhét kiến thức của loài người** mà thôi.

AlphaGo Zero

Tháng 10 năm 2017, AlphaGo chỉ còn là quá khứ. Một phiên bản tân tiến hơn, **AlphaGo Zero đã** ra mắt, mang sức mạnh vượt mặt tất cả các phiên bản tốt nhất của AlphaGo chỉ trong 40 ngày với tỉ số toàn thắng 100-0. Điều đặc biệt chính là, kiến thức của AlphaGo Zero bắt đầu từ **"con số 0", không cần bất cứ nước đi chuyên gia của loài người nữa!**

Vậy **AlphaGo Zero** đã làm gì để thay thế "kiến thức của loài người"? AlphaGo Zero **loại bỏ mạng học có giám sát trên nước đi của chuyên gia**, kết hợp Mạng chính sách và Mạng giá trị thành một mạng nơ-rôn đơn lẻ duy nhất. Đồng thời, cây tìm kiếm Monte Carlo cũng được cải tiến.

"AlphaGo Zero mạnh mẽ đến vậy vì nó không còn bị giới hạn bởi kiến thức của loài người"

Demis Hassabis
– Đồng sáng lập kiêm CEO của DeepMind

AlphaZero

Sau AlphaGo Zero, đội ngũ phát triển DeepMind không hề muốn dừng lại ở mỗi bộ môn cờ vây, mà muốn mở rộng ra nhiều bộ môn với luật chơi và trạng thái biểu diễn đầu vào phức tạp hơn nhiều.

Tháng 12 năm 2017, **AlphaZero** đã được ra mắt để có thể chơi nhiều trò chơi khác như shogi và cờ vua. **AlphaZero** đã dễ dàng đánh bại phần mềm chơi cờ tốt nhất, Stockfish (phần mềm chơi cờ vua mạnh nhất đương thời), Elmo (cờ shogi), và chiến thắng chính phiên bản tiền nhiệm AlphaGo Zero.

Chiến thắng trong cờ vua làm bất ngờ hầu hết các kì thủ hàng đầu, bởi vì người ta luôn cho rằng cờ vua đòi hỏi quá nhiều kiến thức về luật chơi, trạng thái cho một hệ thống học tập tăng cường.

Để lí giải tại sao **AlphaZero** nhanh chóng vượt mặt các thế hệ tiền nhiệm của nó, chúng ta cùng "lặn sâu" vào cây tìm kiếm Monte Carlo, cơ chế tự chơi và các loại bỏ data augmentation được cải tiến như thế nào. Những nội dung này sẽ được đề cập trong những bài viết tiếp theo.

Đôi lời

Sau khi đọc bài viết này, hẳn nhiều người có suy nghĩ về tương lai AI sẽ vượt mặt con người trên mọi lĩnh vực khi chứng kiến những kì thủ vô địch thế giới "câm lặng" trước những nước cờ của máy móc. Song, trong bộ phim tài liệu AlphaGo, có một câu trích dẫn mà tôi rất thích, cũng là lời cảm ơn đến Lee Sedol, khi anh đã một mình đơn độc chiến đấu với đối thủ vô tri này.

"Tôi nghe thấy mọi người đang reo lên vui sướng khi chắc chắn rằng AlphaGo đã thua cuộc. Tôi hiểu rất rõ lí do. Mọi người cảm thấy bất lực và sợ hãi. Có vẻ như con người chúng ta rất yếu ớt và mỏng manh. Và chiến thắng này có nghĩa là chúng ta vẫn có thể giữ vững ngôi vị của mình. Theo thời gian, có lẽ sẽ rất khó để đánh bại AI. Nhưng với một chiến thắng này, cảm giác như thế là đủ. Một lần là đủ."

Tham khảo

[1] "AlphaGo | DeepMind." https://deepmind.com/research/case-studies/alphago-the-story-so-far (accessed Oct. 15, 2020).

[2] "DeepMind's AI became a superhuman chess player in a few hours - The Verge." https://www.theverge.com/2017/12/6/16741106/deepmind-ai-chess-alphazero-shogi-go (accessed Oct. 15, 2020).

[3] "The Evolution of AlphaGo to MuZero | by Connor Shorten | Towards Data Science." https://towardsdatascience.com/the-evolution-of-alphago-to-muzero-c2c37306bf9 (accessed Oct. 15, 2020).

[4] D. Silver *et al.*, "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," Dec. 2017, Accessed: Oct. 15, 2020. [Online]. Available: http://arxiv.org/abs/1712.01815.