Giải pháp SimRL

Cuộc thi: RLCOMP - https://rlcomp.codelearn.io/

Đội: LaoNong

Thứ hạng vòng 1: Hạng 4

I. Giới thiệu

Đội LaoNong chỉ có một thành viên là Lê Tiến Dũng, cựu sinh viên ĐHBKHN, cựu nhân viên FSOFT (thời anh Nam làm giám đốc, anh Đặng Diệu Linh là trưởng dự án). LaoNong tốt nghiệp Tiến Sĩ tại Nhật, hiện đang sinh sống và làm việc tại Vương Quốc Bỉ. Bản thân LaoNong cũng có nhiều kinh nghiệm thi trực tuyến trên Kaggle và đạt danh hiệu Kaggle Competitions Grandmaster.

LaoNong xin cảm ơn BTC cuộc thi đã tạo ra một sân chơi vui vẻ và bổ ích cho cộng đồng AI của Việt Nam. LaoNong cũng xin cảm ơn công ty FPT đã tài trợ cho cuộc thi. Cũng phải đề cập rằng công ty FPT sẽ không sử dụng được giải pháp của các đội vì bài toán cho cuộc thi là một bài toán giả định không có thật. Chính vì điều này mà LaoNong đánh giá rất cao công ty tài trợ và không mang các tiểu xảo "bẩn" áp dụng trong giải pháp của mình.

II. Các hướng tiếp cận

Bài toán đặt ra là bài toán giả định của một trò chơi: Trong trò chơi này, bạn phải điều khiển nhân vật của mình di chuyển trên một bản đồ được thể hiện dưới dạng ma trận hai chiều để vượt qua các chướng ngại vật và đào được nhiều vàng nhất có thể. Hãy chú ý quan sát bản đồ và cẩn thận với thanh năng lượng của mình để đưa ra các chiến thuật đúng đắn. Xin xem thêm chi tiết tại website https://codelearn.io/game/detail/2212875?key=rlcomp#ai-game-summary.

Theo cá nhân LaoNong, có 3 hướng tiếp cận chính cho bài toán

1. Giải thuật

Tức là ta lập trình cho nhân vật (bot) thực hiện mô phỏng các hành động giống suy nghĩ của ta. Cách này tương đối đơn giản, tuy nhiên có có nhược điểm là nếu luật của chúng ta chỉ bao quát một số trường hợp, thì khi gặp trường hợp khác, nhân vật sẽ xử lý sai lệch.

Giải thuật của LaoNong khá đơn giản nhưng giải thuật này dựa trên ý tưởng của nhiều lý thuyết: Optimization, RL (mimic), Social Network Analysis, và Game Theory. Chi tiết của giải thuật được đề cập trong phần sau.

2. Thuần học tăng cường (RL)

RL được xây dựng với đầu vào là thông tin thô (raw) trên bản đồ. Cách này giống Alpha Go, nhưng cách này cần phần cứng tốt, huấn luyện nhân vật với thời gian dài.

3. RL với đặc trưng

Cách này dùng RL nhưng đặc trưng được xây dựng riêng cho bài toán này. Cách xây dựng đặc trưng có thể giống trò chơi PacMan. Bản thân LaoNong ban đầu cũng định dùng DL trên MentalRL (https://github.com/doerlbh/mentalRL). Các đặc trưng sẽ là số vàng tại mỗi mỏ, khoảng cách giữa mỏ vàng / nhân vật ... Tuy vậy cách này vẫn tốn sức lập trình, huấn luyện, rồi thử nghiệm ...

III. Giải thuật SimRL (simulated RL)

Giải thuật SimRL được chia thành nhiều phần nhỏ

1. Tìm đường ngắn nhất từ một vị trí trên bản đồ đến vị trí khác

Trong bài toán này, đường ngắn nhất phải đảm bảo tối ưu về số bước đi và số năng lượng bị mất. Vì mỗi một lần di chuyển thì nhân vật mất đi cơ hội nghỉ nghơi, tương đương với năng lượng khôi phục là 12 đơn vị. Do đó LaoNong đã gán cho chi chí là tổng năng lượng di chuyển cộng 12 đơn vị năng lượng. Sau khi xây dựng được ma trận chi phí di chuyển, thì con đường ngắn nhất là con đường có tổng chi phí nhỏ nhất. Giải thuật này có sẵn trong thư viện scipy với hàm shortest_path.

2. Giá trị kỳ vọng tại mỗi ô vàng nếu như chỉ có một ô vang trên bản đồ

Giá trị này có thể hiểu giống như giá trị kỳ vọng Q của mỗi ô vàng trong RL. Cách tự nhiên nhất thì chính là số vàng tại ô. Tuy nhiên trong RL thì giá trị này bị giảm dần theo thời gian. LaoNong ước lượng giá trị kỳ vọng Q là số vàng * (gamma ** khoảng cách) với gamma là một hằng số, có thể cho gamma là 0.99 chẳng hạn.

3. Giá trị kỳ vọng tại mỗi ô vàng có tham chiếu bởi các ô vàng khác

Vì giá trị kỳ vọng của một ô vàng phải được tính đến giá trị từ các ô vàng bên cạnh. Thế nên tại mỗi ô vàng ta phải cập nhật

Q(s) = alpha * Q(s) + beta * [tổng (Q(s') * gamma ** khoảng cách từ s tới s')]

alpha, beta là các hằng số học: có thể cho alpha là 0.85, beta là 0.15. Tuy nhiên alpha + beta không nhất thiết phải bằng 1.

4. Giá trị kỳ vọng tại mỗi ô vàng có tính tới ảnh hưởng qua lại

Giá trị kỳ vọng có tính chất tương tác, nên ta phải dùng công thức cập nhật trên nhiều lần. LaoNong cho nhân vật cập nhật giá trị Q 3 vòng.

5. Giá trị kỳ vọng tại mỗi ô vàng có tính tới vị trí của các nhân vật khác

LaoNong giả sử rằng các nhân vật khác có suy nghĩ giống nhân vật mình. Như vậy mỗi nhân vật đều có ước lượng Q riêng với

Q(s tại vị trí b) = Q(s) * (gamma ** khoảng cách từ b tới s)

Vì vậy giá trị Q của nhân vật phải là giá trị Q ban đầu trừ giá trị Q của nhân vật khác * hệ số tranh chấp. Trong trường hợp đơn giản ta cho hệ số tranh chấp là 0.25 (cho 4 nhân vật). Tuy nhiên không phải nhân vật nào cũng suy nghĩ giống nhân vật của LaoNong nên hệ số là 80% * 0.25 là 0.2

6. Lựa chọn hành động

Căn cứ vào giá trị Q tại mỗi ô vàng, nhân vật chọn ô vàng có giá trị Q cao nhất rồi đi tới đó. Tất nhiên phải đảm bảo năng lượng hiện thời có thể vượt qua chướng ngại vật ở ô tiếp theo. Nếu không thì nhân vật phải nghỉ nghơi.

IV. Một số tiểu xảo

1. Nghỉ 3 lần

Vì lần nghỉ đầu tiên chỉ khôi phục được 12 năng lượng, còn lần tiếp theo là 16+ cho đến tối đa 50 đơn vị. LaoNong nghỉ luôn 3 lần trừ lúc cuối cuộc thi. Tiểu xảo này không thực sự tối ưu nếu trên đường đi có đầm lầy hoặc nhân vật bị chậm chân khi đến mỏ vòng

2. Di chuyển sớm khỏi mỏ vàng

Việc di chuyển sớm khỏi mỏ vàng sẽ tránh được tranh chấp lúc mỏ vàng gần hết. Tuy nhiên nếu đường đi tiếp theo có bẫy thì cũng không phải tối ưu. Hoặc nếu hết vàng thì việc di chuyển sớm sẽ gây tác dụng ngược lại.

3. Đánh "trung lộ"

Do LaoNong nhận thấy giải thuật bị lỗi tối ưu cục bộ, nhân vật có thể di chuyển vào góc và không kịp quay ra. Đội Beerbot thì cho phép nhân vật "lật cánh đánh đầu". Tuy nhiên kỹ thuật này khó nên LaoNong cho nhân vật đánh trung lộ rồi mới chọn góc đánh.

Tiểu xảo này có đòn phản tác dụng trong mấy lượt thi đấu ở vòng loại số 2 dẫn tới đội LaoNong bị dừng chân.

V. Kết luận

Giải pháp SimRL đã đạt được hạng 4 ở vòng loại sở khảo (đấu với nhận vật của BTC) và vượt qua vòng loại số 1 cũng phần nào thể hiện tính hiệu quả của giải thuật. Việc áp dụng tiểu xảo có tính 2 mặt và rủi ro cao so với các giải thuật có sự hậu thuẫn của các lý thuyết đã được kiểm chứng.

LaoNong hy vọng có dịp tiếp tục tham gia sân chơi ở các cuộc thi tiêp theo.