**Áp dụng giải thuật học tăng cường cho thị trường mua bán cổ phiếu**

by

Bùi Quang Minh

FPT University

[minhbqhe170952@fpt.edu.vn](mailto:minhbqhe170952@fpt.edu.vn)

1. **Đặt vấn đề**

Áp dụng học tăng cường vào lĩnh vực mua bán cổ phiếu đang trở thành một xu hướng nổi bật trong thế giới tài chính hiện đại. Học tăng cường, một nhánh của trí tuệ nhân tạo, không chỉ cho phép máy tính học hỏi từ kinh nghiệm mà còn tối ưu hóa các quyết định dựa trên môi trường biến động và phức tạp của thị trường chứng khoán. Thông qua việc sử dụng các thuật toán học tăng cường, nhà đầu tư có thể xây dựng các mô hình dự đoán và chiến lược giao dịch có khả năng thích ứng và cải thiện hiệu quả qua thời gian. Điều này đặc biệt quan trọng trong bối cảnh thị trường chứng khoán luôn biến động mạnh và khó lường. Tuy nhiên, việc áp dụng học tăng cường vào mua bán cổ phiếu cũng đặt ra nhiều thách thức, từ việc thu thập và xử lý dữ liệu lớn đến việc đảm bảo tính ổn định và tránh rủi ro trong các quyết định tự động. Vì vậy, nghiên cứu và triển khai học tăng cường trong mua bán cổ phiếu không chỉ mở ra những cơ hội lớn mà còn đòi hỏi sự cẩn trọng và chuyên sâu để khai thác tối đa tiềm năng của công nghệ này.

1. **Mô hình**
   1. **Giải thuật**

A2C (Advantage Actor Critic) là một thuật toán học tăng cường kết hợp giữa các phương pháp Actor-Critic và đồng bộ hóa để đào tạo các mô hình học sâu. Nguyên lý hoạt động của A2C có thể được giải thích như sau:

* Actor: Là một mô hình mạng nơ-ron chịu trách nhiệm quyết định hành động nào sẽ thực hiện dựa trên trạng thái hiện tại của môi trường. Actor dự đoán một phân phối xác suất cho các hành động có thể được thực hiện.
* Critic: Là một mô hình mạng nơ-ron khác đánh giá hành động mà Actor đã thực hiện. Critic ước tính giá trị của trạng thái hiện tại (Value Function) hoặc giá trị của cặp trạng thái-hành động (Q-value).
  1. **Action space**

Có hai hành động có thể thực hiện:

* Sell (0): Hành động này đại diện cho việc bán cổ phiếu.
* Buy (1): Hành động này đại diện cho việc mua cổ phiếu.

Bên cạnh đó chúng ta cũng có 2 positions: (Long position) muốn mua cổ phiếu khi giá thấp và kiếm lời bằng cách giữ chúng trong khi giá trị của chúng đang tăng lên, và vị thế bán khống (Short position) muốn bán cổ phiếu với giá cao và sử dụng số tiền đó để mua lại cổ phiếu ở mức giá thấp hơn, giữ lại phần chênh lệch làm lợi nhuận.

A black screen with white text

Description automatically generated

* 1. **Observation space**

Hàm observation giúp agent có cái nhìn liên tục về thị trường trong một khoảng thời gian ngắn hạn tùy thuộc vào giới hạn của tham số window\_size. Giả sử nếu như window size là 5, điều này có nghĩa là agent sẽ được quan sát biến động của cổ phiếu và các features trong vòng 5 ngày liên tục.



A screenshot of a computer program

Description automatically generated

* 1. **Reward**

Hàm reward tính toán phần thưởng dựa trên sự thay đổi giá cổ phiếu và vị thế hiện tại của tác nhân. Nếu tác nhân đang ở vị thế Long và giá cổ phiếu tăng so với lần giao dịch cuối cùng, tác nhân sẽ nhận được phần thưởng dương tương ứng với sự chênh lệch giá. Điều này khuyến khích tác nhân mua khi giá thấp và giữ cho đến khi giá tăng để tối đa hóa lợi nhuận.

A screen shot of a computer code

Description automatically generated

1. **Kết quả**

Số vốn ban đầu là : 100.000. Dữ liệu được train trong vòng 100000 timesteps, giao dịch từ 2014 – 2024

Kết quả: 1.606.751 , reward: 1165

A screen shot of a computer

Description automatically generated

A green line graph on a white background

Description automatically generated

**Bonus: kết quả test 1 năm gần nhất**

Kết quả: 206.099

A computer screen with text

Description automatically generated