


MẪU BÁO CÁO CỦA MỖI HV

Họ và tên (IN HOA)	TRẦN QUỐC THÀNH (CH2001035) LƯƠNG MẠNH HÀ (CH2001028)
Ảnh	
Số buổi vắng	0
Bonus	30
Tên đề tài (VN)	MÔ HÌNH GIAO DỊCH CHỨNG KHOÁN TỰ ĐỘNG SỬ DỤNG PHƯƠNG PHÁP HỌC TĂNG CƯỜNG VÀ XỬ LÝ NGÔN NGỮ TỰ NHIÊN
Tên đề tài (EN)	STOCK DAY TRADING MODEL USING REINFORCEMENT LEARNING AND NATURAL LANGUAGE PROCESSING
Giới thiệu	<i>Những năm gần đây, phong trào đầu tư tài chính mà cụ thể là đầu tư chứng khoán phát triển mạnh mẽ ở Việt Nam. Số lượng tài khoản mở mới trung bình hằng tháng đã tăng từ mức 20,000 năm 2017 lên mức trên 110,000 năm 2021. Thanh khoản trung bình của hệ thống cũng tăng từ mức 4,000 tỷ đồng mỗi phiên giao dịch năm 2017 lên mức 30,000 tỷ đồng năm 2021. Tổng vốn hóa những công ty trên sàn chứng khoán đến tháng 5 năm 2021 đã đạt đến gần 8 triệu tỷ đồng, gấp 1.3 lần GDP (tổng sản phẩm quốc nội) [1]. Những năm sắp tới, thị trường chứng khoán được dự báo bùng nổ hơn nữa vì tỉ lệ người dân đầu tư chứng khoán của Việt Nam vẫn còn rất nhỏ so với những nước phát triển. Tỉ lệ này ở Việt Nam là 3% (tháng 5 năm 2021) so với mức gần 50% ở một số nước phát triển. Như vậy, đầu tư chứng</i>

khoán là lĩnh vực còn rất nhiều tiềm năng và sẽ phát triển mạnh mẽ ở Việt Nam trong thời gian tới.

Một trong những đặc điểm của việc đầu tư chứng khoán là giá của các cổ phiếu thay đổi rất nhanh và phức tạp. Việc hàng triệu nhà đầu tư cùng lúc ra quyết định mua bán dựa vào nhiều nguồn thông tin khác nhau (báo điện tử, diễn đàn mạng xã hội) dẫn đến việc dự đoán giá cổ phiếu trong ngắn hạn là rất khó khăn. Các nhà đầu tư thường sử dụng một số phân tích kỹ thuật (dựa vào giá của cổ phiếu đó trong quá khứ) hoặc phân tích cơ bản (sử dụng thông tin từ báo cáo tài chính của doanh nghiệp và một số thông tin vĩ mô về ngành). Phân tích kỹ thuật có nhược điểm là bỏ qua các thông tin về doanh nghiệp, các yếu tố vĩ mô còn phân tích cơ bản thiếu thông tin về giá trong quá khứ và chậm hơn thị trường (báo cáo tài chính thường đi sau hoạt động thực tế của doanh nghiệp vài tháng).

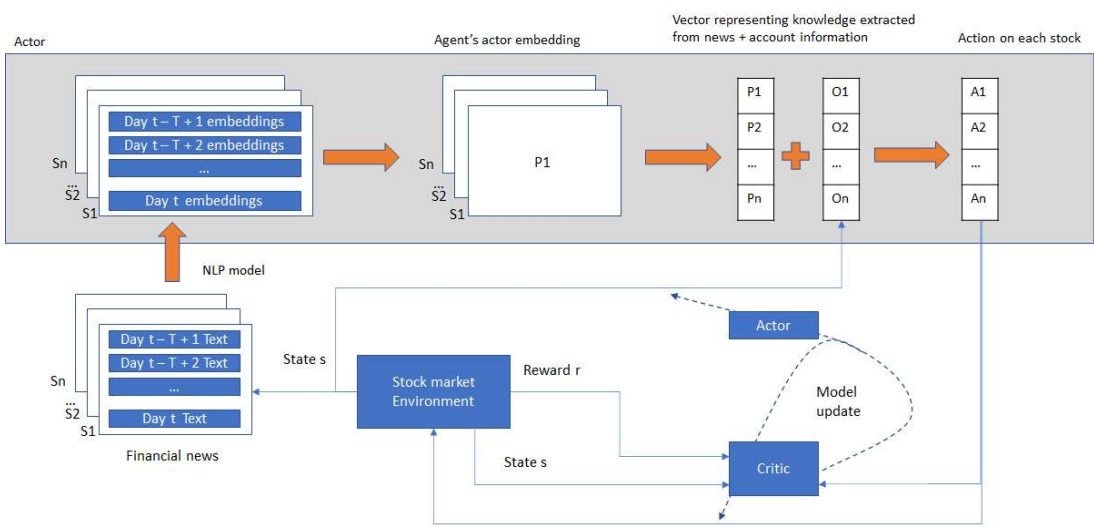
Trước tình hình đó, những nhà đầu tư tổ chức và các công ty chứng khoán đã xây dựng các hệ thống mua bán tự động dựa vào 1 số chỉ báo. Tuy nhiên các chỉ báo này thường rất thô sơ, chủ yếu so sánh giá cổ phiếu hiện tại với một giá cố định được đặt trước hoặc dựa vào giá cổ phiếu trong quá khứ (ví dụ chỉ số sức mạnh tương đối RSI, đường trung bình động MA, dải Bollinger BB, ...). Đặc điểm của những chỉ số này là chỉ quan tâm đến giá của một cổ phiếu cụ thể trong quá khứ mà không kết hợp thông tin từ việc thay đổi giá của các cổ phiếu khác và thông tin từ báo điện tử và mạng xã hội. Hệ thống mua bán tự động từ đó cũng chỉ ra quyết định mua bán độc lập mà không ra quyết định dựa trên tổng thể các mã cổ phiếu và số tiền hiện tại (quản lý danh mục - portfolio management), chủ yếu được dùng trong các trường hợp cực đoan (giá cổ phiếu giảm quá sâu trong thời gian ngắn).

Về mặt học thuật, các mô hình học sâu với đầu vào là giá chứng khoán quá khứ và thông tin ngôn ngữ tự nhiên từ báo chí và mạng xã hội đã đạt được các kết quả vượt trội so với các mô hình học máy cổ điển trong việc dự đoán giá cổ phiếu trong tương lai [2]. Tuy nhiên, từ dự đoán giá cổ phiếu đến tự động mua bán cổ phiếu có rất nhiều vấn đề cần nghiên cứu, ví dụ số vốn hiện có, các mã chứng khoán hiện có trong tài khoản, chi phí giao dịch,... Phương pháp học tăng cường đã chứng minh được hiệu quả trong việc ra quyết định trong thời gian ngắn để đạt được lợi ích tối đa trong dài hạn (điều mà các phương pháp dự đoán giá cổ phiếu không

có), có tính đến yếu tố chi phí giao dịch, danh mục tiền mặt và chứng khoán hiện có, ... rất phù hợp trong ứng dụng đầu tư chứng khoán tự động. Ramit Sawhney và Arnav Wadhwa đã ứng dụng phương pháp học tăng cường để thực nghiệm giao dịch trên sàn chứng khoán New York (NASDAQ), Shanghai, Shenzhen, Hong Kong [3] với kết quả vượt trội so với các mô hình học sâu khác trong quá khứ.

Tại Việt Nam, một vài nghiên cứu về dự đoán giá chứng khoán trên sàn HOSE đã được thực hiện trong thời gian gần đây, chủ yếu sử dụng mô hình học sâu (LSTM) với đầu vào là dữ liệu về giá chứng khoán trong quá khứ [4]. Tất cả các nghiên cứu này đều thiếu các yếu tố để có thể ra quyết định giao dịch chứng khoán tự động như: không có đánh giá hiệu quả lâu dài, không đề cập đến chi phí giao dịch, không đề cập đến số vốn ban đầu, ...

Mục tiêu của đề tài này là xây dựng môi trường chứng khoán ảo với thông tin tài chính và giá chứng khoán thu thập từ quá khứ để phục vụ việc nghiên cứu giao dịch tự động, từ đó sử dụng và đánh giá một số phương pháp học tăng cường đang được phát triển hiện nay.



Hình 1: mô hình học tăng cường cho tự động giao dịch chứng khoán

Mục tiêu Mục tiêu tổng quát: xây dựng mô hình và đánh giá hệ thống giao dịch tự động trên thị trường chứng khoán Việt Nam với mục tiêu lợi nhuận trong trung và dài hạn

	<p>(trên 3 tháng) cao hơn lãi suất ngân hàng (6%/năm) và tăng trưởng của chỉ số VN-Index trong cùng thời gian.</p> <p>Mục tiêu cụ thể:</p> <ul style="list-style-type: none"> • Hiểu rõ và lập lại được các phương pháp đề cập trong tài liệu tham khảo [2] và [3] về tự động giao dịch sử dụng học tăng cường . • Tìm kiếm và thu thập dữ liệu của tin tức và giá cổ phiếu ở Việt Nam trong quá khứ với tổng thời gian ít nhất 2 năm liên tục. Từ đó xây dựng môi trường giao dịch ảo sử dụng dữ liệu thu thập được. • Xây dựng hệ thống giao dịch tự động trong môi trường giao dịch ảo ứng dụng phương pháp học tăng cường với mục tiêu tối đa lợi nhuận. Mục tiêu lợi nhuận đạt được lớn hơn chỉ số VN-Index và lãi suất ngân hàng trong thời kiểm tra gian tối thiểu 2 tháng. • Thử nghiệm hệ thống với giao dịch tự động trực tiếp trên sàn HOSE (mục tiêu phụ, trong điều kiện thời gian, kinh phí cho phép và khả năng hợp tác với công ty chứng khoán).
<p>Nội dung và phương pháp thực hiện</p>	<p>Đề tài nghiên cứu bao gồm những nội dung sau:</p> <ul style="list-style-type: none"> • Đọc và hiểu rõ các tài liệu trong phần tham khảo. • Tập hợp dữ liệu liên quan từ các tài liệu tham khảo. • Xây dựng code dựa vào tài liệu tham khảo hoặc dựa vào code của các tài liệu tham khảo (nếu có sẵn). • Đánh giá lại các phương pháp trong phần tài liệu tham khảo. • Thu thập dữ liệu từ các trang báo điện tử liên quan đến lĩnh vực chứng khoán (https://cafef.vn, https://tinnhanhchungkhoan.vn/, https://vietstock.vn/, https://ndh.vn/, https://www.thesaigontimes.vn/). • Thu thập dữ liệu giá chứng khoán và khối lượng giao dịch (nguồn https://vn.investing.com) • Xây dựng môi trường giao dịch chứng khoán ảo dựa vào dữ liệu tin tức và giá chứng khoán thu thập được, môi trường phải theo tiêu chuẩn gym-openai: https://gym.openai.com)

	<ul style="list-style-type: none"> • Huấn luyện agent theo phương pháp học tăng cường trong môi trường chứng khoán ảo, đánh giá kết quả đạt được. • Kết nối agent đã xây dựng với môi trường giao dịch chứng khoán thực tế.
Kết quả dự kiến	<ul style="list-style-type: none"> • Dữ liệu tin tức tài chính từ báo điện tử và giá chứng khoán được phân loại theo mã chứng khoán và sắp xếp theo thứ tự thời gian. • Môi trường giao dịch chứng khoán ảo dùng để huấn luyện agent theo phương pháp học tăng cường. • Ứng dụng giao dịch chứng khoán tự động với lợi nhuận lớn hơn sự tăng trưởng của chỉ số VN-Index và lãi suất ngân hàng.
Tài liệu tham khảo	<p>[1] Ủy ban chứng khoán nhà nước - https://www.scc.gov.vn/ (7/2021)</p> <p>[2] Weiwei Jiang: Applications of deep learning in stock market prediction: recent progress. CoRR abs/2003.01859 (2020)</p> <p>[3] Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, Rajiv Ratn Shah: Quantitative Day Trading from Natural Language using Reinforcement Learning. NAACL-HLT 2021: 4018-4030</p> <p>[4] Huyen Giang Thi Thu, Thuy Nguyen Thanh, Tai Le Quy: A Neighborhood Deep Neural Network Model using Sliding Window for Stock Price Prediction. BigComp 2021: 69-74</p>