



DỰ BÁO GIÁ CỔ PHIẾU TRÊN THỊ TRƯỜNG CHỨNG KHOÁN THÔNG QUA MÔ HÌNH PHỨC HỢP LSTM – GRU HYBRID

TRINH VIẾT GIANG

Giá cổ phiếu là dữ liệu phi tuyến rất phức tạp, bị ảnh hưởng bởi nhiều yếu tố, do đó dự đoán chỉ số giá cổ phiếu là công việc rất khó khăn. Tại Việt Nam, hiện rất ít mô hình phức hợp sử dụng máy học và học sâu được thiết kế bởi Python với các gói hỗ trợ có sẵn để dự báo biến số kinh tế. Nghiên cứu này đánh giá so sánh các mô hình học sâu LSTM, GRU và mô hình phức hợp của chúng để dự báo cho chỉ số Vn-Index trên Sở Giao dịch Chứng khoán Hồ Chí Minh từ ngày 01/01/2009 đến ngày 01/01/2024. Kết quả đánh giá mô hình học sâu phức hợp cho thấy tỷ lệ sai lệch so với thực tế là thấp nhất, từ đó, khuyến nghị các doanh nghiệp, các nhà đầu tư và nhà hoạch định chính sách nên tăng cường sử dụng công nghệ trong việc dự báo nhằm hỗ trợ đưa ra quyết định.

Từ khóa: Dự báo, Giá cổ phiếu, HOSE, LSTM, Gru, LSTM-GRU Hybrid, Vn-Index

FORECASTING STOCK PRICES ON THE STOCK MARKET THROUGH LSTM – GRU HYBRID COMPLEX MODEL

Trinh Viet Giang

Stock prices are highly complex, nonlinear data, influenced by numerous factors, making the prediction of stock price indices a challenging task. In Vietnam, there are currently very few complex models using machine learning and deep learning, designed with Python and its available support packages, for forecasting economic variables. This study focuses on a comparative evaluation of deep learning models, including LSTM, GRU, and their hybrid models, to forecast the Vn-Index on the Ho Chi Minh Stock Exchange from January 1, 2009, to January 1, 2024. The evaluation results of the hybrid deep learning model show the lowest deviation from actual values. Hence, it is recommended that businesses, investors, and policymakers should enhance the use of technology in forecasting to support decision-making processes.

Keywords: Forecast, stock prices, HOSE, LSTM, Gru, LSTM-GRU Hybrid, Vn-Index

Ngày nhận bài: 9/9/2024

Ngày hoàn thiện biên tập: 18/9/2024

Ngày duyệt đăng: 25/9/2024

Giới thiệu

Dự đoán giá cổ phiếu là đối tượng nghiên cứu của nhiều học giả với các phương pháp dự đoán và mô hình hóa chỉ số giá cổ phiếu như: mô hình truyền

thống và các mô hình mạng nơ-ron. Thực tế đã cho thấy các mô hình truyền thống không thể nắm bắt được các mẫu phi tuyến tính. Để khắc phục hạn chế này, các mô hình mạng nơ-ron đã được sử dụng rộng rãi trong việc dự đoán các chuỗi thời gian phi tuyến tính (Ferreira và cộng sự, 2008) như chỉ số giá cổ phiếu.

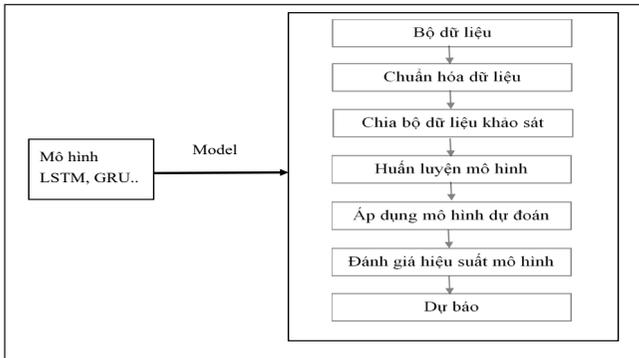
Nghiên cứu này sử dụng kết hợp giữa LSTM, GRU để phát huy thế mạnh của các mô hình dự báo như: mạng LSTM có dung lượng bộ nhớ dài hơn để lưu giữ và xử lý thông tin trước đó, do đó, đối với dữ liệu lớn, mạng LSTM có thể mang lại kết quả tốt hơn. Tuy nhiên, GRU nhanh hơn nhiều so với LSTM vì nó có ít tham số hơn. Để tận dụng lợi thế của mỗi mô hình, bài viết này đã kết hợp LSTM và GRU, và đề xuất một mô hình mạng GRU-LSTM chính quy mới có hiệu suất tốt hơn giúp dự đoán giá đóng cửa của chỉ số Vn-Index.

Cơ sở lý thuyết và phương pháp nghiên cứu

Cơ sở lý thuyết

Chỉ số VN-Index: Là chỉ số đại diện cho toàn bộ cổ phiếu niêm yết tại Sở Giao dịch Chứng khoán TP. Hồ Chí Minh (HOSE). Được tính toán dựa trên giá và khối lượng giao dịch của các cổ phiếu trên sàn HOSE, VN-Index phản ánh mức độ biến động chung của TTCK Việt Nam. Đây là một trong những chỉ số quan trọng nhất cho nhà đầu tư và người theo dõi TTCK. VN-Index thể hiện xu hướng giá cổ phiếu hàng ngày thông qua việc so sánh giá trị vốn hoá

HÌNH 1: PHƯƠNG PHÁP DỰ BÁO THEO MÔ HÌNH HỌC SÂU



Nguồn: Tác giả tổng hợp

của thị trường ở hiện tại với giá trị vốn hoá của thị trường cơ sở, xem chúng đã thay đổi bao nhiêu lần.

Mô hình nghiên cứu

Mô hình LSTM

Hochreiter (1997) lần đầu tiên đề xuất LSTM vào năm 1997 và sau đó đã trở nên phổ biến rộng rãi, đặc biệt là để sử dụng trong việc giải quyết các vấn đề liên quan đến dự đoán chuỗi thời gian. LSTM bổ sung cấu trúc "cổng" so với Mạng nơ-ron hồi quy (RNN) thông thường, mỗi ô nhớ thường có ba cổng: cổng vào, cổng quên và cổng ra. Chức năng chính của các cổng vào là tiếp nhận một loạt thông tin và xác định thông tin nào có thể được giữ lại trong trạng thái ô; cổng quên quyết định loại thông tin nào sẽ bị loại bỏ khỏi trạng thái ô, bước này có thể đạt được bằng hàm kích hoạt sigmoid. Kết quả tính toán và giá trị đầu ra có thể thu được thông qua cổng ra.

Mô hình GRU

Mô hình GRU hay còn gọi là mạng nơ-ron hồi tiếp với nút cổng, được giới thiệu bởi Cho và cộng sự (2014). GRU có thể xem là một biến thể thế hệ mới của LSTM, về cấu trúc GRU đơn giản chỉ với 02 cổng: Cổng khởi tạo (reset gate) và cổng cập nhật (update gate). Với thiết kế đơn giản như vậy, mô hình GRU giảm số lượng tham số so với LSTM, điều này làm cho mô hình này nhanh hơn trong việc huấn luyện và dự báo.

Mô hình GRU-LSTM phức hợp

Mô hình phức hợp của nghiên cứu này là sử dụng 02 lớp ẩn đầu tiên của mô hình LSTM hoặc GRU làm 02 lớp ẩn đầu vào của một mô hình tổng quát. Bài nghiên cứu này sử dụng thuật ngữ LSTM-GRU Hybrid chỉ cách sử dụng 02 lớp ẩn của LSTM làm đầu vào của

mô hình và GRULSTM Hybrid là trường hợp ngược lại.

Các chỉ tiêu đánh giá chất lượng dự báo

- Sai số tuyệt đối trung bình (MAE): Là dạng thước đo hồi quy đơn giản nhất, nó được tính bằng cách lấy giá trị tuyệt đối trung bình của các phần dư của giá trị thực tế và giá trị dự đoán. Mỗi phần dư sẽ tỷ lệ với tổng sai số, nghĩa là các sai số lớn cũng sẽ tuyến tính với tổng sai số. Một MAE nhỏ sẽ gọi ý rằng mô hình rất tốt trong việc dự đoán.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

- Sai số bình phương trung bình (MSE): MSE chỉ đơn giản đề cập đến giá trị trung bình của chênh lệch bình phương giữa tham số dự đoán và tham số quan sát được. MSE sẽ luôn lớn hơn MAE nên không thể so sánh MSE với MAE. Tuy nhiên, có thể so sánh các giá trị này với các chỉ số của một mô hình khác. Điều này sẽ giúp chúng ta chọn mô hình tốt hơn cho dữ liệu.

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

- Căn bậc hai của sai số bình phương trung bình (RMSE): Lỗi trung bình bình phương gốc (RMSE) là thước đo mức độ hiệu quả của mô hình. RMSE càng nhỏ tức là sai số càng bé thì mức độ ước lượng cho thấy độ tin cậy của mô hình có thể đạt cao nhất.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

- Sai số tương đối trung bình (MAPE): Phản ánh giá trị dự báo sai khác bao nhiêu phần trăm so với giá trị trung bình.

$$MAPE = \frac{1}{n} \sum_{k=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

HÌNH 2: HƯỚNG BIẾN ĐỘNG CHỈ SỐ VN-INDEX TỪ NGÀY 01/1/2009 ĐẾN NGÀY 01/01/2024



Nguồn: Phần mềm Python



Trong đó:

+ n: Tổng số mẫu

+ y_i : Giá trị thực tế mẫu thứ i

+ \hat{y}_i : Giá trị dự đoán mẫu thứ i

- Đánh giá mức độ phù hợp của mô hình R2 (r² score - Rsquared score): Là một chỉ số quan trọng để đánh giá hiệu quả của các mô hình hồi quy, nhưng nên được sử dụng kết hợp với các chỉ số khác để có cái nhìn toàn diện hơn về hiệu suất của mô hình. Giá trị r² score nằm trong khoảng từ 0 đến 1 và cho kết quả gần tương đương nhau, nếu giá trị của nó càng gần 1 thì mô hình dự báo càng hiệu quả:

$$r_{core}^2 = \frac{\sum_{i=1}^n |y_i - \bar{y}|^2 - \sum_{i=1}^n |y_i - \hat{y}_i|^2}{\sum_{i=1}^n |y_i - \bar{y}|^2}$$

Trong đó:

+ n: Tổng số mẫu

+ y_i : Giá trị thực tế mẫu thứ i

+ \hat{y}_i : Giá trị dự đoán mẫu thứ i

+ \bar{y} : Giá trị trung bình của mẫu i

Phương pháp nghiên cứu và dữ liệu nghiên cứu

Phương pháp nghiên cứu

Phương pháp nghiên cứu của bài viết được minh họa qua Hình 1.

Dữ liệu nghiên cứu

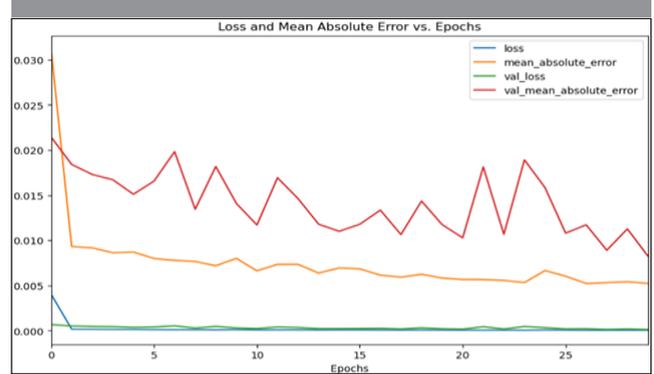
Bài viết này chọn giá trị chỉ số Vn-Index giao dịch lúc đóng cửa hàng ngày của HOSE làm biến số ủy quyền của giá giao dịch chứng khoán làm bộ dữ liệu nghiên cứu chính. Toàn bộ cả dữ liệu được lấy từ cơ sở dữ liệu của M.S Fusion Media Ltd (<https://www.investing.com/indices/vn-historicaldata>) từ ngày 01/01/2009 đến ngày 01/01/2024 với tổng số 3.753 quan sát. Trong mô hình học sâu, bộ dữ liệu huấn luyện chứa 3.002 quan sát, chiếm 80% tập dữ liệu gốc và bộ dữ liệu giám sát chứa 751 quan sát, chiếm 20% cuối cùng của tập dữ liệu gốc, cụ thể biểu đồ đặc tả xu hướng biến động chỉ số Vn-Index trước - Hình 2(a) và sau khi chia tập dữ liệu chính - Hình 2(b).

BẢNG 1: ĐO LƯỜNG ĐỘ CHÍNH XÁC TRÊN TẬP GIÁM SÁT CỦA 3 MÔ HÌNH LSTM, GRU, LSTM-GRU

Thông số	LSTM	GRU	LSTM-GRU Hybrid
RMSE	20.224	18.385	17.1823
MAE	15.150	13.773	12.8728
MSE	409.016	338.079	262.2162
MAPE	0.01554	0.01149	0.0098
R2 Score	0.9716	0.9796	0.9876

Nguồn: Tác giả tổng hợp

HÌNH 3: CÁC CHỈ SỐ MẤT MẮT (LOSSES) CỦA MÔ HÌNH



Nguồn: Phần mềm Python

Kết quả nghiên cứu

Trong bài viết này, các mô hình đơn lẻ và phức hợp LSTM, GRU được triển khai bằng cách sử dụng các gói của ngôn ngữ lập trình Python trong Jupyter Notebook.

Kết quả chạy mô hình LSTM trong Python

Mô hình này bao gồm ba lớp LSTM (có kích thước 150, 100, và 100 đơn vị) kết hợp với các lớp Dropout nhằm ngăn ngừa overfitting. Sau đó là ba lớp Dense để chuyển đổi dữ liệu từ LSTM thành các đầu ra dự báo. Mô hình này có 279.129 tham số và không có tham số nào không thể huấn luyện. Mô hình LSTM này được thiết kế để dự báo chuỗi thời gian với nhiều đầu ra.

Kết quả chạy GRU trong Python

Mô hình GRU với 128 đơn vị GRU và 4 đầu ra là một kiến trúc tương đối nhỏ và gọn. Số lượng tham số chỉ hơn 51.000 giúp mô hình dễ dàng huấn luyện và tính toán, đặc biệt khi tài nguyên phần cứng hạn chế. Với cấu trúc đơn giản, mô hình này phù hợp với các bài toán dự đoán chuỗi thời gian hoặc bài toán có đầu ra nhỏ, không quá phức tạp.

Kết quả chạy mô hình phức hợp LSTM-GRU trong Python

Mô hình LSTM-GRU Hybrid là một cách tiếp cận mạnh mẽ để xử lý chuỗi thời gian, tận dụng điểm mạnh của cả LSTM và GRU để ghi nhớ thông tin dài hạn và ngắn hạn. Mô hình có 167.684 tham số, giúp cân bằng giữa độ phức tạp và hiệu quả trong các bài toán dự báo chuỗi thời gian phức tạp hoặc đa biến.

Kết quả đo lường các chỉ số MAPE và R2 Score ở Bảng 1 cho thấy hai mô hình đều cho độ phù hợp khá tốt. Tuy nhiên, mô hình LSTM-GRU Hybrid giá trị dự đoán chỉ số Vn-Index thấp nhất so với giá trị thực tế (1,84%) và sự phù hợp của mô hình là 98,16%. Do đó, mô hình LSTM-GRU Hybrid có chỉ số đo

lượng thống kê tốt nhất, làm cơ sở cho tiến hành các dự báo tiếp theo trong nghiên cứu.

Các chỉ số mất mát của mô hình theo các epoch trong quá trình huấn luyện LSTM-GRU Hybrid được thể hiện ở Hình 3.

Nghiên cứu này áp dụng mô hình dự báo xu hướng biến động của chỉ số Vn-Index với thời gian dự báo liên tiếp là 60 ngày đầu tiên và 60 ngày kế tiếp. Kết quả cho thấy chỉ số Vn-Index tăng nhanh năm 2012-2022 và giảm mạnh năm 2023.

Thảo luận kết quả nghiên cứu

Nghiên cứu đã cho thấy mô hình LSTM-GRU Hybrid là sự lựa chọn tối ưu bằng cách đánh giá hiệu suất của mô hình với các chỉ tiêu: MSE, RMSE, MAE, MAPE, R2 Score cho dự báo giá chứng khoán. Kết quả nghiên cứu cũng phù hợp với các nghiên cứu trước về mô hình phức hợp đối với chuỗi thời gian tài chính (Alpaydin, 2020; Patra & Mohanty, 2022; Trần Đăng, 2024). Mô hình LSTM-GRU phức hợp không chỉ kế thừa các ưu điểm của từng mô hình đơn lẻ mà còn khắc phục được những hạn chế của chúng. Điều này càng khẳng định thêm rằng mô hình LSTM-GRU Hybrid là lựa chọn thích hợp cho nhiệm vụ dự báo chỉ số chứng khoán trên TTCK Việt Nam trong thời gian tới.

Kết luận và khuyến nghị

Trên cơ sở kết quả nghiên cứu, tác giả đề xuất một số khuyến nghị như sau:

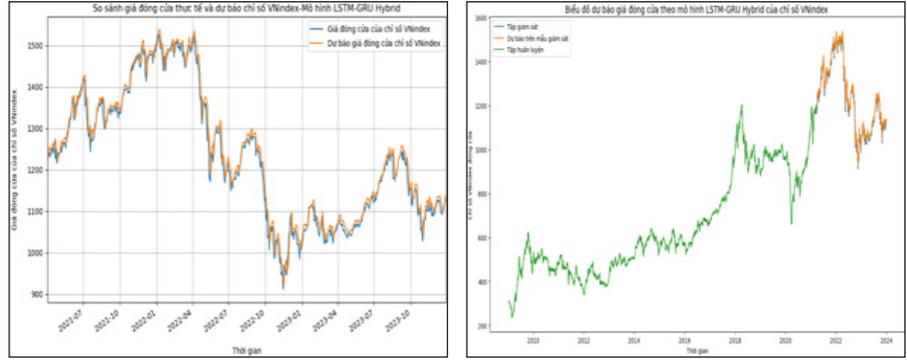
Đối với nhà đầu tư, doanh nghiệp

Bên cạnh các phương pháp phân tích truyền thống như phương pháp phân tích cơ bản trên sàn chứng khoán, cần tìm hiểu học tập nghiên cứu nhằm ứng dụng các phương pháp hiện đại dựa vào học máy, học sâu LSTM-GRU Hybrid để dự báo giá chứng khoán đem lại độ chính xác cao. Từ đó, nhà quản trị đưa ra quyết định mua, bán cổ phiếu, phát hành cổ phiếu, tăng vốn hay tái cấu trúc tài chính. Hơn nữa, việc sử dụng dự báo giúp nhà quản trị quản lý tài sản và dòng tiền tốt hơn, tối ưu hóa lợi nhuận từ các khoản đầu tư.

Đối với cơ quan nhà nước

Nhà nước có thể sử dụng các mô hình dự báo để nhận diện sớm các dấu hiệu bất ổn trên thị

HÌNH 4: SO SÁNH KẾT QUẢ DỰ BÁO TRÊN TẬP DỮ LIỆU HUẤN LUYỆN VÀ GIÁM SÁT CỦA CHỈ SỐ VN-INDEX



Nguồn: Phần mềm Python

trường, như bong bóng tài sản hoặc sự sụt giảm đột ngột của một số ngành nhất định. Điều này giúp cơ quan quản lý can thiệp sớm bằng các biện pháp chính sách thích hợp, ngăn ngừa các rủi ro hệ thống. Bên cạnh đó, Nhà nước cũng cần có các chính sách hỗ trợ phát triển hạ tầng công nghệ và dữ liệu tài chính; Tăng cường tính minh bạch trong việc chia sẻ dữ liệu thị trường nhằm đảm bảo tính chính xác của các mô hình dự báo. Ngoài ra, Nhà nước có thể xây dựng khung pháp lý hỗ trợ cho các doanh nghiệp công nghệ tài chính, đồng thời tạo điều kiện cho việc áp dụng học sâu trong dự báo tài chính một cách an toàn và hiệu quả.

Tài liệu tham khảo:

1. Đào Lê Kiều, O., & Nguyễn Thị Minh, C. (2024), Dự báo chỉ số chứng khoán bằng học máy: Bằng chứng thực nghiệm từ TTCK Việt Nam. *Tạp chí Kinh tế và Dự báo*, 1, 6;
2. Dương Ngân, H. (2018), Dự báo biến động của chỉ số VN-Index thông qua khối lượng giao dịch ròng và giá trị giao dịch ròng của nhà đầu tư nước ngoài. *Tạp chí Khoa học & Đào tạo Ngân hàng*, 195, 18-25;
3. Trần Đăng, T. (2024), Đánh giá hiệu suất mô hình phức hợp LSTM-GRU: nghiên cứu điển hình về dự báo chỉ số đo lường xu hướng biến động giá cổ phiếu trên Sàn Giao dịch Chứng khoán TP. Hồ Chí Minh. *Tạp chí Khoa học Đại học Cần Thơ*, 60(1), 235-249;
4. Alpaydin, E. (2020), *Introduction to machine learning*: MIT press;
5. Liu, Y., Wang, Z., & Zheng, B. (2019), Application of regularized GRU-LSTM model in stock price prediction. Paper presented at the 2019 IEEE 5th International Conference on Computer and Communications (ICCC);
6. Patra, G. R., & Mohanty, M. N. (2022), An LSTM-GRU based hybrid framework for secured stock price prediction. *Journal of Statistics and Management Systems*, 25(6), 1491-1499.

Thông tin tác giả:

TS. Trịnh Viết Giang
 Trường Đại học Công nghiệp Hà Nội
 Email: trinhvietgiang@hau.edu.vn