

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**NIÊN LUẬN CƠ SỞ
NGÀNH CÔNG NGHỆ THÔNG TIN**

Đề tài

**PHƯƠNG PHÁP DỰ ĐOÁN GIÁ CỔ PHIẾU VỚI
ECHO STATE NETWORKS**

**Sinh viên: Lê Quang Sang
Mã số: B1606927
Khóa: 42**

Cần Thơ, 11/2019

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG
BỘ MÔN CÔNG NGHỆ THÔNG TIN**

**NIÊN LUẬN CƠ SỞ
NGÀNH CÔNG NGHỆ THÔNG TIN**

Đề tài

**PHƯƠNG PHÁP DỰ ĐOÁN GIÁ CỔ PHIẾU VỚI
ECHO STATE NETWORKS**

**Người hướng dẫn
TS Lâm Nhật Khang**

**Sinh viên thực hiện
Lê Quang Sang
Mã số: B1606927**

Khóa: K42

Cần Thơ, 11/2019

Lời cảm ơn

Cảm ơn TS. Lâm Nhựt Khang, Bộ môn Công nghệ Thông tin, Khoa Công nghệ Thông tin và Truyền thông, trường Đại học Cần Thơ đã tích cực hướng dẫn, giúp đỡ trong nghiên cứu đề tài này.

Mục lục

Tóm lược	1
Phần giới thiệu.....	2
Phần nội dung:.....	3
Chương 1 - Đặc tả yêu cầu:	3
Chương 2 - Thiết kế giải pháp.....	3
1. Lý do chọn chuỗi thời gian (Time series).....	3
2. Các thuộc tính của Time series	3
3. Tại sao chúng ta quan tâm về tính dừng của dữ liệu:.....	5
4. Xác định tính dừng của dữ liệu:.....	5
5. Phương pháp dự đoán chuỗi thời gian cơ bản với Echo State Network:	5
Chương 3 – Cài đặt giải pháp	6
Chương 4. Đánh giá kiểm thử	7
Phần kết luận	10
Tài liệu tham khảo	11

Danh mục hình ảnh và biểu bảng

Hình 1 Giá trị trung bình không đổi	4
Hình 2 Giá trị phương sai không đổi	4
Hình 3 Tính tương tự không phụ thuộc vào thời gian	5
Hình 4 Echo state networks	6
Hình 5 Mô hình hóa dữ tập dữ liệu	7
Hình 6 Biểu đồ nhiệt của hàm tính toán lỗi	7
Hình 7 Kết quả dự đoán mô hình tổng thể	8
Hình 8 Kết quả dự đoán của 100 điểm dữ liệu	8
Hình 9 Kết quả dự đoán cho 10 ngày tiếp theo	9
Hình 10 Biểu đồ thể hiện lỗi của cửa sổ trượt	10

Tóm lược

Với sự phát triển nhanh chóng của khoa học công nghệ ngày nay, máy tính có thể làm rất nhiều công việc khó khăn, đòi hỏi sự tính toán và suy luận logic thay cho con người, trong đó bao gồm các công việc dự đoán. Trong đó có việc dự đoán giá cổ phiếu trên thị trường chứng khoán, hiện có nhiều nghiên cứu thực hiện việc nghiên cứu dự báo giá của cổ phiếu trên thị trường chứng khoán với các phương pháp và thuật toán khác nhau. Trong đó phương pháp điển hình nhất và được nghiên cứu sử dụng gần đây là dùng mạng Echo State Networks (ESN), hiện cho kết quả rất khả quan.

Phần giới thiệu

Dự đoán sự biến động của giá cổ phiếu trên thị trường chứng khoán đang là một vấn đề đang được các nhà đầu tư quan tâm, làm sao để các nhà đầu tư có thể có một quyết định đúng đắn nhất trong việc lựa chọn hay không lựa chọn đầu tư trên thị trường chứng khoán ngày nay. Theo những nghiên cứu gần đây thì Echo State Networks cho chúng ta kết quả dự đoán khả quan nhất.

Mục tiêu của đề tài là tìm hiểu cách hoạt động của của mạng Echo State Networks trong việc dự đoán giá cổ phiếu cho 2 ngày tiếp theo dựa trên những dữ liệu có sẵn của những ngày trước đó.

Bố cục của bản báo cáo gồm 3 phần: Phần giới thiệu, Phần nội dung và Phần kết luận. Trong đó Phần nội dung gồm có 4 chương:

Chương 1 - Đặc tả yêu cầu: Giới thiệu phương pháp dự đoán giá cổ phiếu với Echo State Networks.

Chương 2 - Thiết kế giải pháp: Trình bày những kiến thức liên quan đến Echo State Networks.

Chương 3 - Cài đặt giải pháp: Mô tả quá trình hoạt động của phương pháp dự đoán giá cổ phiếu với Echo State Networks .

Chương 4 - Đánh giá kiểm thử: Trình bày mục tiêu, phương pháp và kết quả kiểm thử.

Phần nội dung:

Chương 1 - Đặc tả yêu cầu:

Phương pháp dự đoán giá cổ phiếu với Echo State Networks được trình bày trong đề tài này dựa trên phương pháp được trình bày trong nghiên cứu “Echo State Network” của Herbert Jaeger^[1] (Jacobs University Bremen, Bremen, Germany) đăng trong Scholarpedia.

Giá cổ phiếu trên thị trường chứng khoán thường được đưa về bài toán là chuỗi thời gian (time series). Các nghiên cứu ngày nay thường sử dụng phương pháp stochastic (ngẫu nhiên) hoặc cải tiến của stochastic (ngẫu nhiên) ví dụ mô hình ARIMA, RegARIMA, . . . để đưa ra các giá trị dự đoán phù hợp với các giá trị trong quá khứ. Mục tiêu cuối cùng là để phản ánh quy luật của thị trường và từ đó sinh ra lợi nhuận.

Chương 2 - Thiết kế giải pháp

Phương pháp dự đoán giá cổ phiếu được đưa về bài toán chuỗi thời gian (time series).

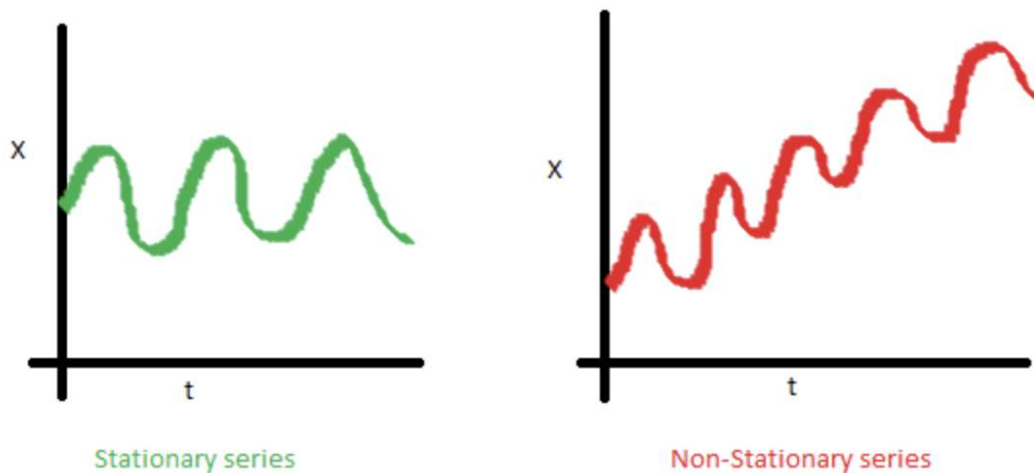
1. Lý do chọn chuỗi thời gian (Time series)

Thị trường chứng khoán được xem là chuỗi thời gian hỗn loạn. Các phương pháp dự đoán ngẫu nhiên được nhiều công ty áp dụng để đưa ra dự đoán chính xác nhất có thể và để họ có thêm lợi thế trên thị trường, vấn đề ở đây là lựa chọn một chuỗi thời gian dự đoán đáng tin cậy.

2. Các thuộc tính của Time series

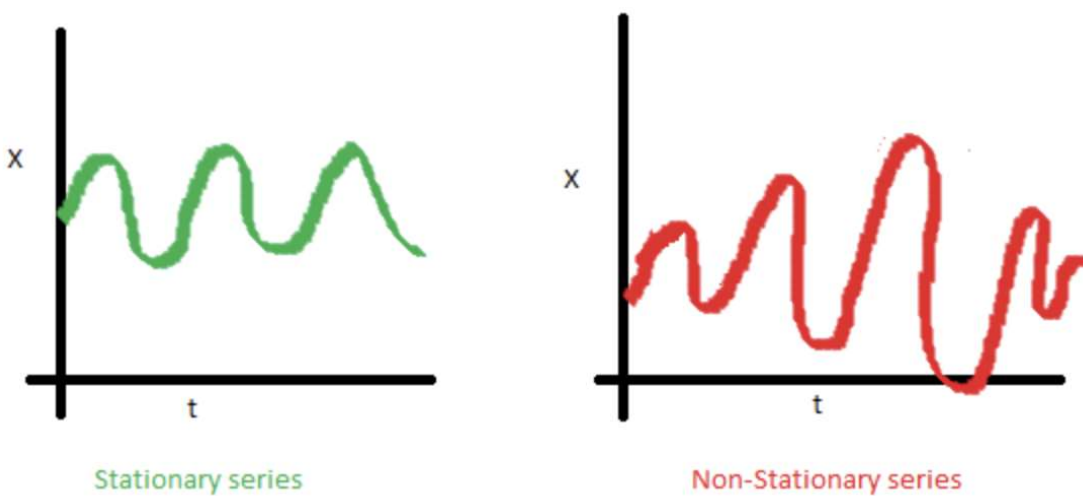
Một trong các thuộc tính của chuỗi thời gian là tính dừng (Stationary). Một time series có tính dừng nếu các thuộc tính thống kê của nó (vd: trung bình, độ lệch chuẩn,...) không đổi theo thời gian. Đa số các mô hình về chuỗi thời gian được xây dựng dựa trên giả định tính dừng của chuỗi thời gian. Nếu chuỗi thời gian ở trong quá khứ có một hành vi nào đó thì khả năng cao là nó sẽ lặp lại trong tương lai. Ngoài ra để hoàn thiện về tính dừng của chuỗi thời gian và cũng đã có một số nghiên cứu lý thuyết về tính dừng của dữ liệu sẽ dễ thao tác, tính toán hơn so với dữ liệu không dừng. Tuy nhiên, trong bài toán thực tế, chúng ta có thể giả định rằng một chuỗi time series được coi là có tính dừng nếu các thuộc tính thống kê không đổi theo thời gian, nghĩa là:

- Giá trị trung bình không thay đổi. Nếu giá trị trung bình thay đổi, chuỗi thời gian sẽ có khuynh hướng đi lên hoặc đi xuống



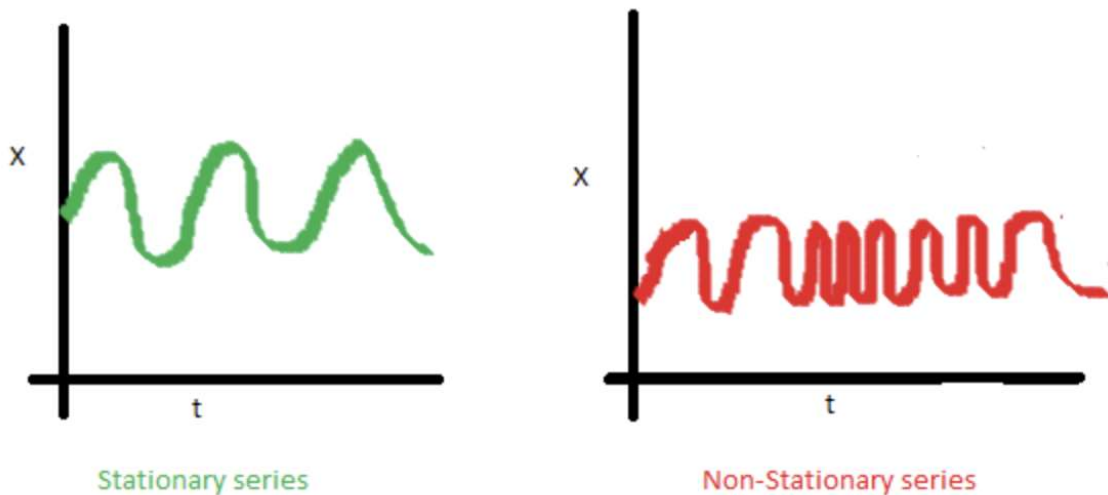
Hình 1 Giá trị trung bình không đổi

- Giá trị phương sai không thay đổi. Thuộc tính này còn được gọi là đồng đẳng (homoscedasticity). Hình bên dưới mô tả một chuỗi có phương sai thay đổi (không có tính dừng) và một chuỗi có phương sai bất biến (có tính dừng).



Hình 2 Giá trị phương sai không đổi

- Tính tương tự không phụ thuộc vào thời gian:



Hình 3 Tính tương tự không phụ thuộc vào thời gian

3. Tại sao chúng ta quan tâm về tính dừng của dữ liệu:

Nếu dữ liệu không có tính dừng thì chúng ta không thể xây dựng mô hình chuỗi thời gian như đã đề cập ở trên. Trong trường hợp nếu chúng ta có dữ liệu mà vi phạm 1 trong 3 tính chất ở trên của tính dừng dữ liệu thì chúng ta cũng có một số phương pháp để đưa dữ liệu về dữ liệu có tính dừng là khử xu hướng (detrending) và khử sai biệt (differencing).

4. Xác định tính dừng của dữ liệu:

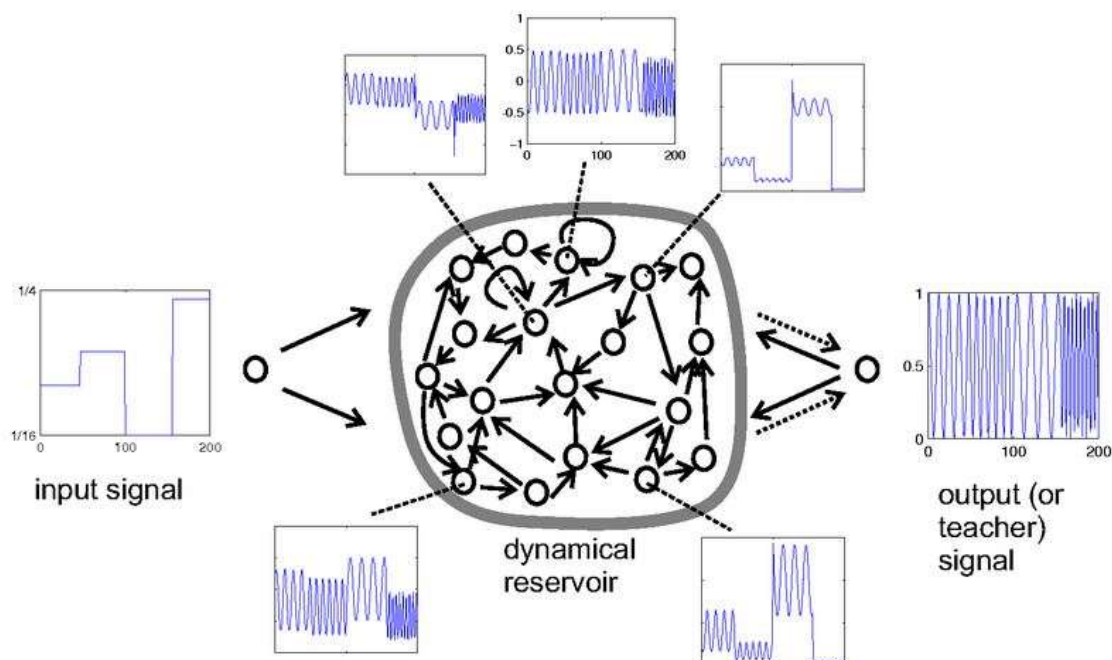
Để xác định một biểu đồ chuỗi time series có tính dừng hay không (quan sát biểu đồ bằng mắt). Cho nên chúng ta sẽ sử dụng kiểm định Dickey-Fuller^[2]. Đây là một kiểm định thống kê để kiểm tra xem chuỗi dữ liệu có tính dừng hay không. Với giả thuyết null là chuỗi time series là một chuỗi không có tính dừng. Nếu giá trị nhỏ hơn một ngưỡng p-value nào đó (thường là 0.05), chúng ta có quyền bác bỏ giả định null, và nói rằng chuỗi thời gian đang có là có tính dừng.

5. Phương pháp dự đoán chuỗi thời gian cơ bản với Echo State Network:

Echo state networks (ESN)^[3] với ý tưởng tương tự như một mạng hồi quy (RNN)^[4] là sửa lỗi các đầu vào (input) tới các kết nối ẩn (hidden connection) và các kết nối ẩn với nhau (hidden -> hidden connections) tại các giá trị khởi tạo ngẫu nhiên và chỉ học từ lớp ẩn cuối cùng tới đầu ra (output), với các lớp ẩn kết nối thưa thớt (thường là 1% nối kết). Trọng số đầu vào được gán cố định ngẫu nhiên và chúng được mở rộng, chúng ta chỉ huấn luyện lớp đầu ra cuối cùng với một mô hình tuyến tính. Khả năng tương đồng với SVM (support vector machine) một thuật toán học có giám sát để phân cụm dữ liệu.

ESN nhận vào giá trị thực tế thay đổi theo thời gian với đầu vào mong muốn là một sóng hình sin và đầu ra có thể học được cũng là một sóng hình sin chỉ định bởi lớp đầu vào và huấn luyện với mô hình tuyến tính các trạng thái của các đơn vị ẩn. Trung tâm

là một hồ chứa động (Dynamical reservoir) với các đơn vị ẩn kết nối với nhau đầu ra(output) cũng có thể trở lại làm một đơn vị ẩn trong hồ chứa, các đơn vị ẩn (hidden unit) giao động rất gần với nhau tạo nên các kết nối ẩn.

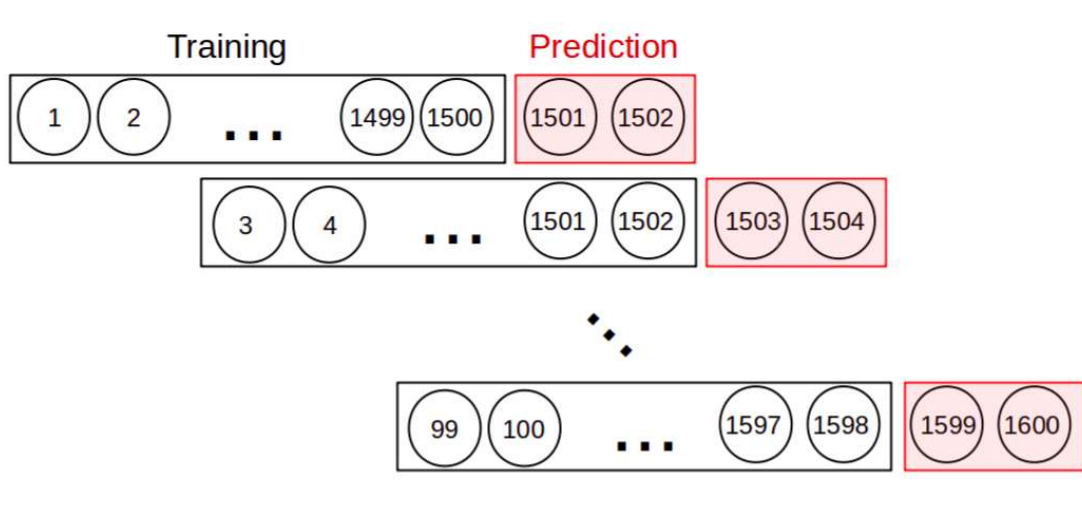


Hình 4 Echo state networks

Về cơ bản chúng ta chỉ đào tạo (training) trọng số đầu ra, điều này giúp tăng tốc độ đào tạo mô hình so với cách mạng thần kinh nhân tạo truyền thống, đó là ưu điểm của ESN (việc tính toán hồ chứa).

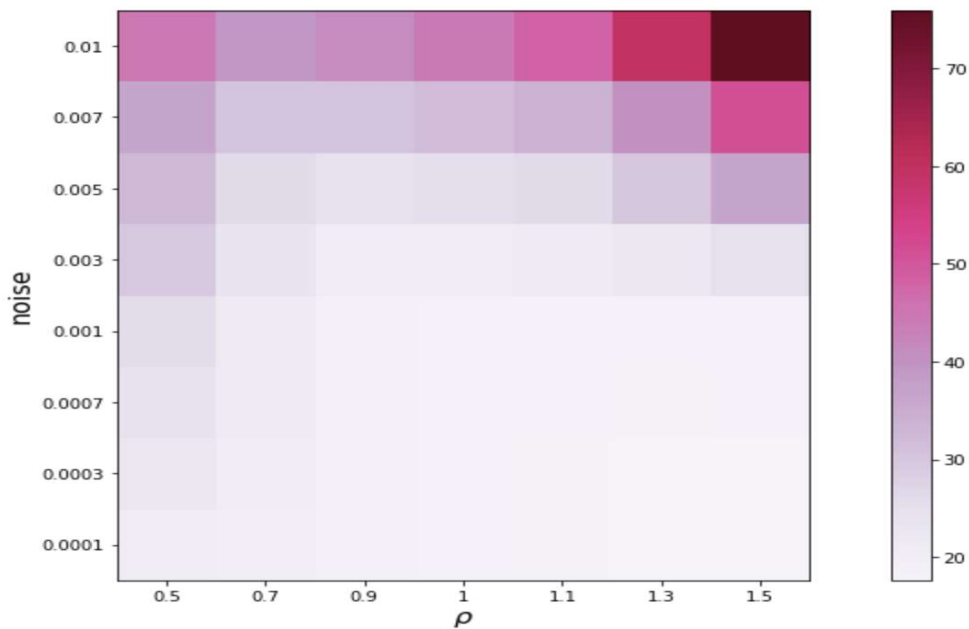
Chương 3 – Cài đặt giải pháp

Chúng ta sẽ dùng Python để cài đặt cho toàn bộ chương trình với thư viện pyESN^[6]. Nhiệm vụ của chúng ta là dự đoán giá cổ phiếu của 2 ngày kế tiếp bằng cách sử dụng 1500 điểm dữ liệu trước đó và thực hiện điều này cho 100 điểm dữ liệu trong tương lai.



Hình 5 Mô hình hóa dữ tập dữ liệu

Đầu tiên chúng ta hình thành một mạng ESN với các tham số hợp lý, sau đó chúng ta tạo một hàm tính toán lỗi MSE^[5] (Độ lệch bình phương trung bình), khi vẽ MSE trên bản đồ nhiệt chúng ta nhận được kết quả sau đây.

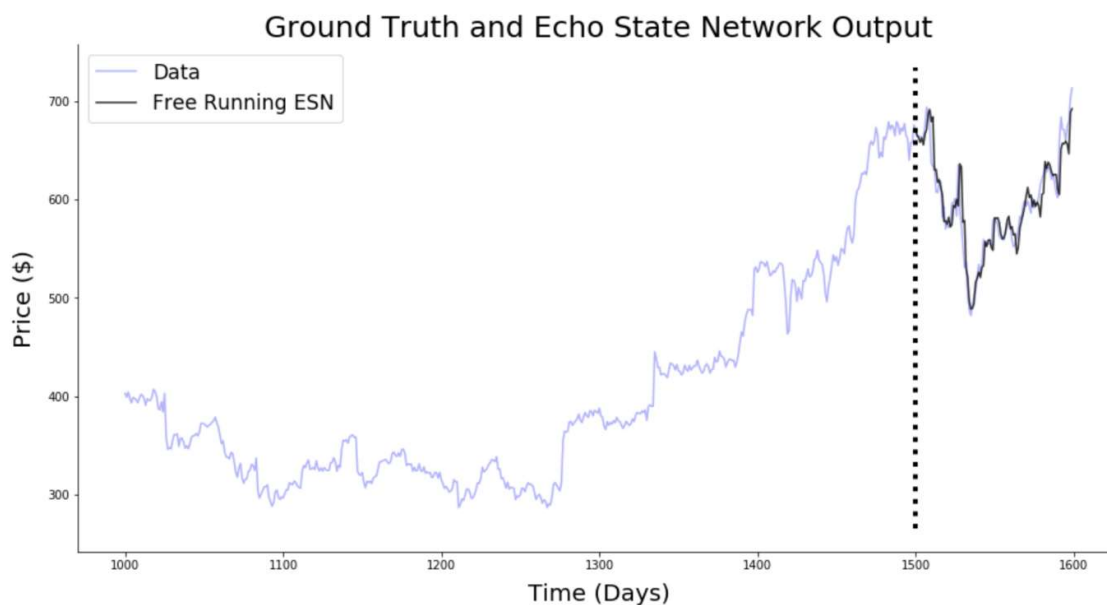


Hình 6 Biểu đồ nhiệt của hàm tính toán lỗi

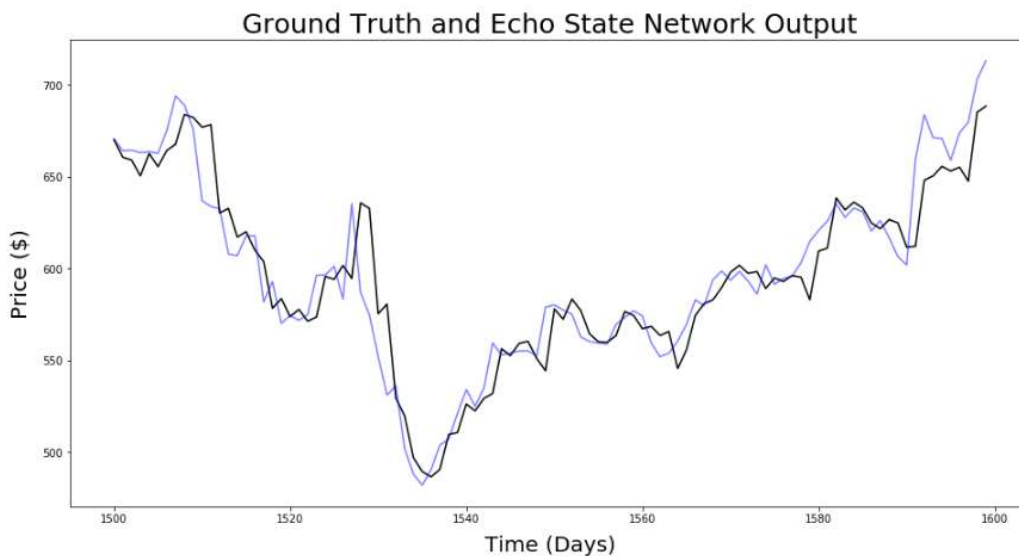
Chương 4. Đánh giá kiểm thử

Sử dụng tập dữ liệu giá cổ phiếu 1500 ngày của Amazon^[7] tiến hành đánh giá ta thu được kết quả khá khả quan của mô hình.

Kết quả thu được:

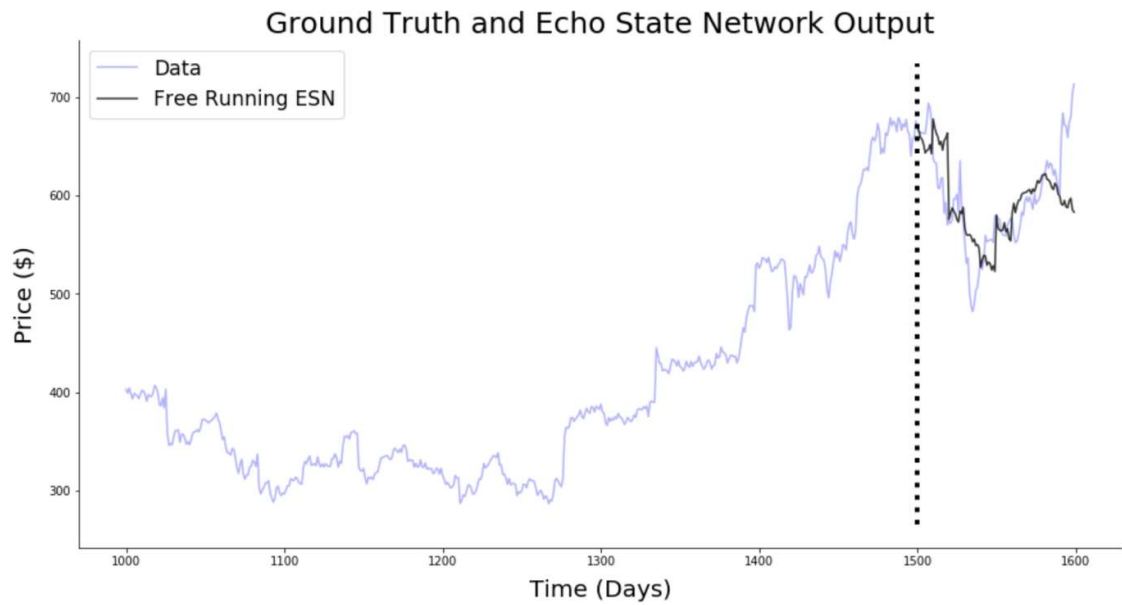


Hình 7 Kết quả dự đoán mô hình tổng thể



Hình 8 Kết quả dự đoán của 100 điểm dữ liệu

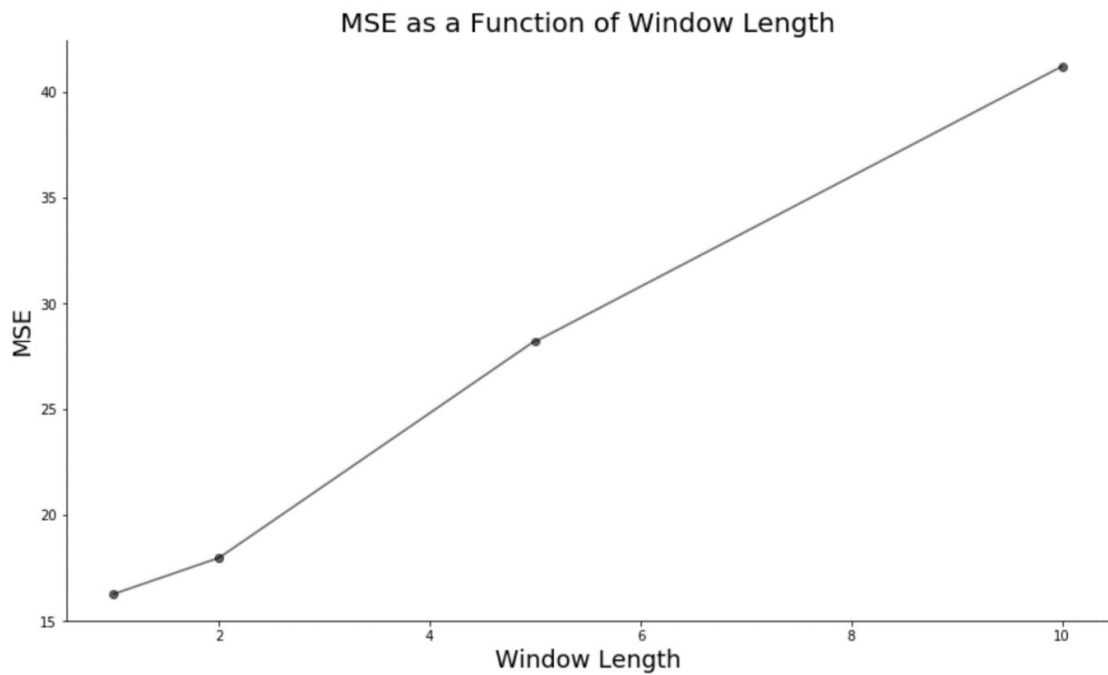
Và sau đây chúng ta thử dự toán cho 10 ngày tiếp theo sau đây là hình ảnh minh họa khi chúng ta tăng kích thước cửa sổ lên 10, kết quả khá tệ so với kích thước bằng 2.



Hình 9 Kết quả dự đoán cho 10 ngày tiếp theo

Bài học rút ra ở đây là mô hình dường như hoạt động tốt với kích thước cửa sổ trượt nhỏ (1 hoặc 2) với độ chính xác hợp lý. Nhưng các lỗi xuất hiện ngày càng nhiều khi chúng ta tăng kích thước cửa sổ trượt lên.

Trong dự đoán tương lai, lỗi lan truyền theo thời gian đây chính là lý do mà độ chính xác của mô hình giảm theo thời gian khi tăng chiều dài cửa sổ trượt lên. Chúng ta có thể thấy ở hình bên dưới MSE tăng khi tăng chiều dài của cửa sổ trượt.



Hình 10 Biểu đồ thể hiện lỗi của cửa sổ trượt

Phần kết luận

Khả năng của mạng Echo State Network (ENS) là để phân tích một chuỗi thời gian hỗn loạn làm cho nó trở thành một công cụ hữu ích cho việc dự đoán tài chính, giá cổ phiếu trên thị trường chứng khoán nơi mà dữ liệu vô cùng hỗn loạn và phi tuyến tính. Ngoài ra chúng ta còn có thể ứng dụng Echo State Network nhiều hơn trong các lĩnh vực:

1. Dự báo thời tiết
2. Thực hiện nhận dạng mẫu
3. Điều khiển các hệ thống động lực học phức tạp

Tài liệu tham khảo

1. Herbert Jaeger. Jacobs University Bremen, Bremen, Germany. Echo State Network. Scholarpedia
2. Dickey-Fullter https://en.wikipedia.org/wiki/Dickey%E2%80%93Fuller_test
3. Echo State Networks (ESN) https://en.wikipedia.org/wiki/Echo_state_network
4. Recurrent neural networks (RNN) https://en.wikipedia.org/wiki/Recurrent_neural_network
5. MSE (Mean Squared Error) https://en.wikipedia.org/wiki/Mean_squared_error
6. pyESN Library <https://github.com/cknd/pyESN/blob/master/pyESN.py>
7. Amazon dataset <https://github.com/littleguppies/Study/blob/master/jupyter-notebook/EchoStateNetworks/amazon.txt>

