

ỨNG DỤNG MẠNG NEURON NHÂN TẠO TRONG VIỆC DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN CÓ TÍNH XU HƯỚNG VÀ TÍNH MÙA

Luận văn tốt nghiệp

GVHD: PGS.TS Dương Tuấn Anh

SVTH 1: Đoàn Ngọc Bảo 50800107

SVTH 2: Ngô Duy Khánh Vy 50802706

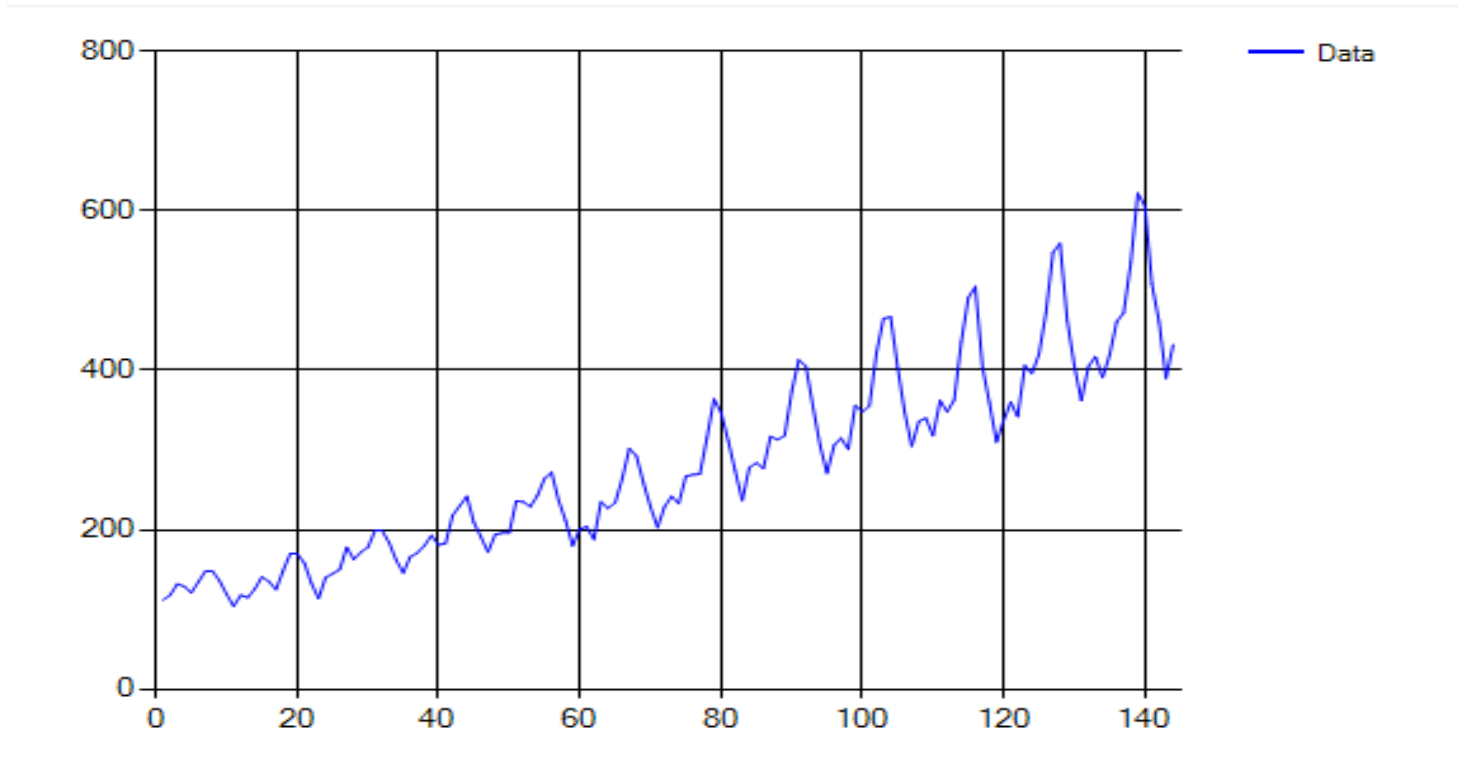
Nội dung

- Đặt vấn đề
- Mạng neuron nhân tạo truyền thẳng
- Áp dụng mạng neuron vào dự báo dữ liệu chuỗi thời gian
- Mô hình lai
- Mô hình khử mùa, khử xu hướng
- Thực nghiệm
- Kết luận
- Hướng phát triển
- Q&A

Đặt vấn đề

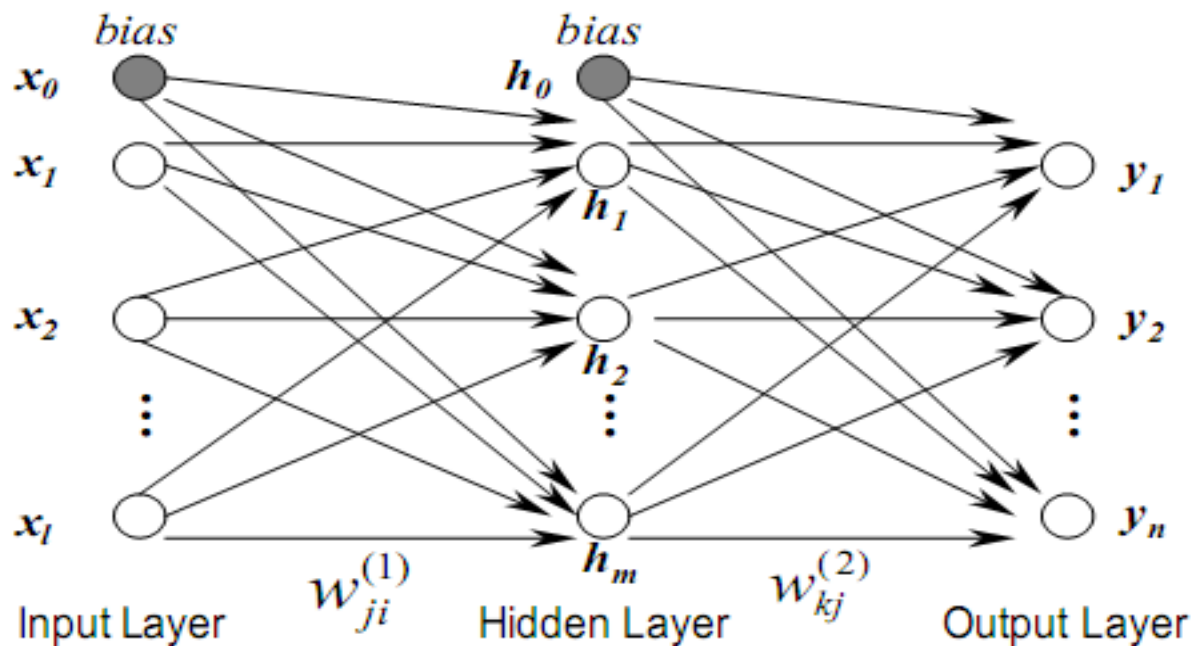
- Dữ liệu chuỗi thời gian là một dạng dữ liệu được quan sát và lưu trữ theo thời gian.
- Một chuỗi thời gian được biểu diễn bằng một chuỗi số thực $\{y_t\}$, với t là biến thời gian.
- Một chuỗi thời gian mà giá trị của nó tăng hay giảm dần trong một giai đoạn dài hạn là chuỗi có tính xu hướng.
- Một chuỗi thời gian mà giá trị của biến đổi lặp đi lặp lại theo từng thời điểm cố định trong năm là chuỗi thời gian có tính mùa.

Đặt vấn đề



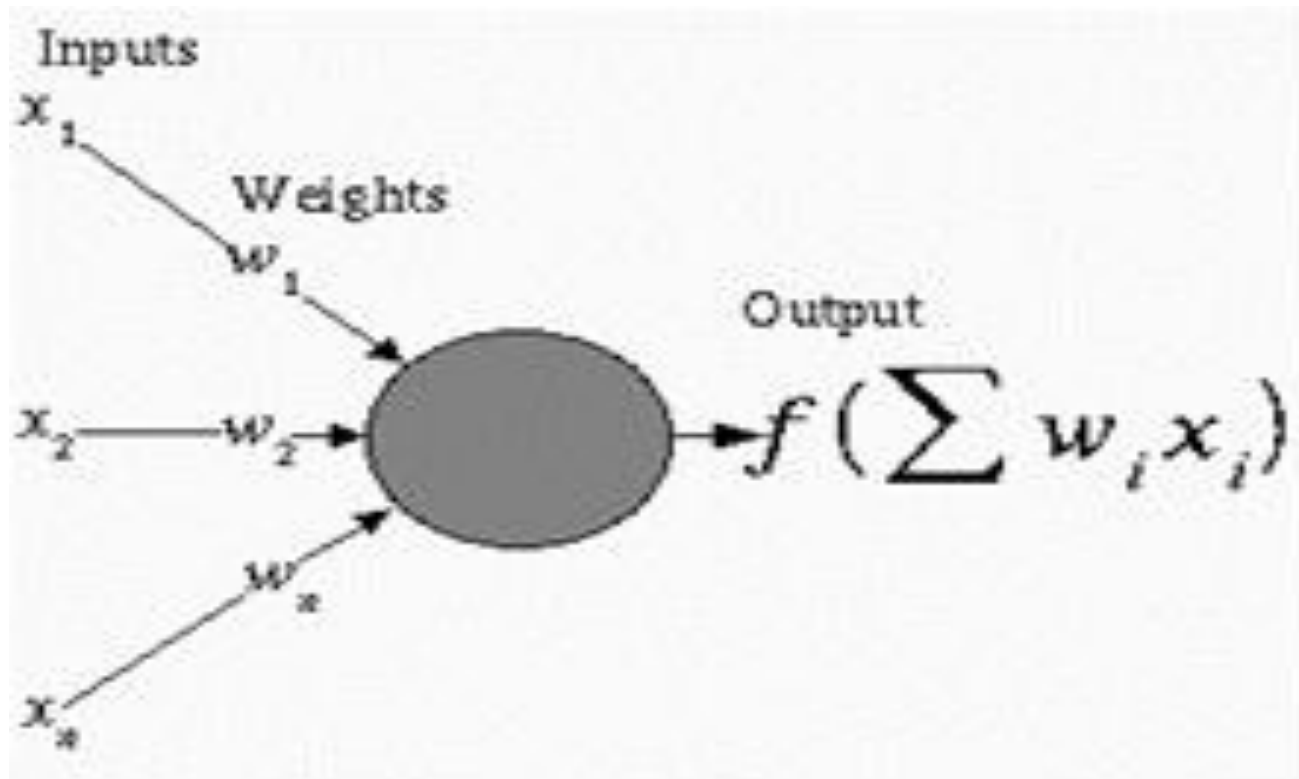
Lượng khách hàng đặt chỗ hàng tháng của hãng hàng không Pan Am từ năm 1946 đến năm 1960

Mạng Neuron nhân tạo truyền thẳng



Mạng neuron truyền thẳng

Mạng Neuron nhân tạo truyền thẳng



Đơn vị mạng neuron

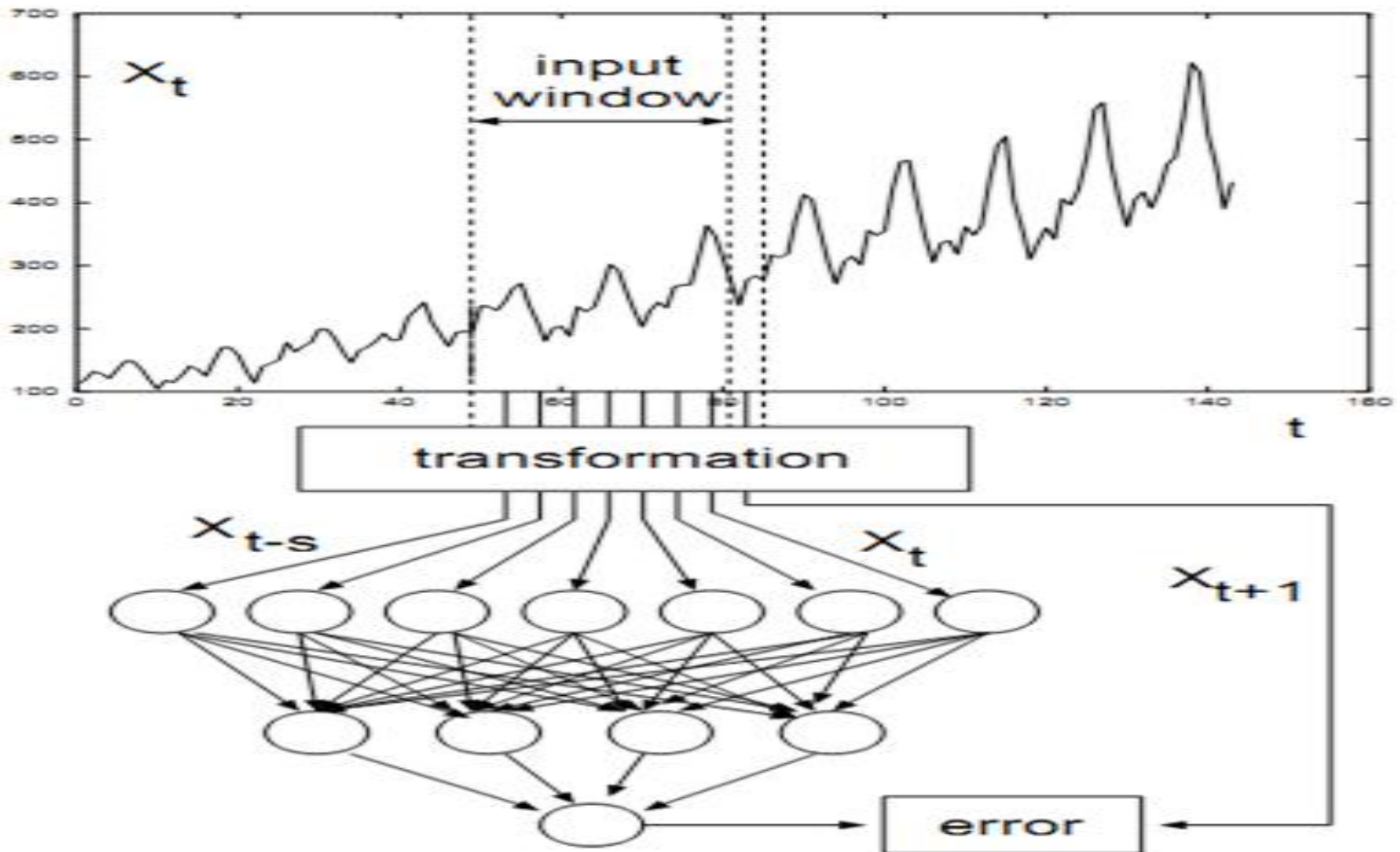
Mạng Neuron nhân tạo truyền thẳng

- Để mạng sinh ra được đầu ra mong muốn phải xác định đúng các trọng số w_{ij} bằng một giải thuật huấn luyện
- Trong luận văn này, chúng tôi đã hiện thực và sử dụng hai giải thuật huấn luyện sau:
 - Giải thuật lan truyền ngược
 - Giải thuật RPROP

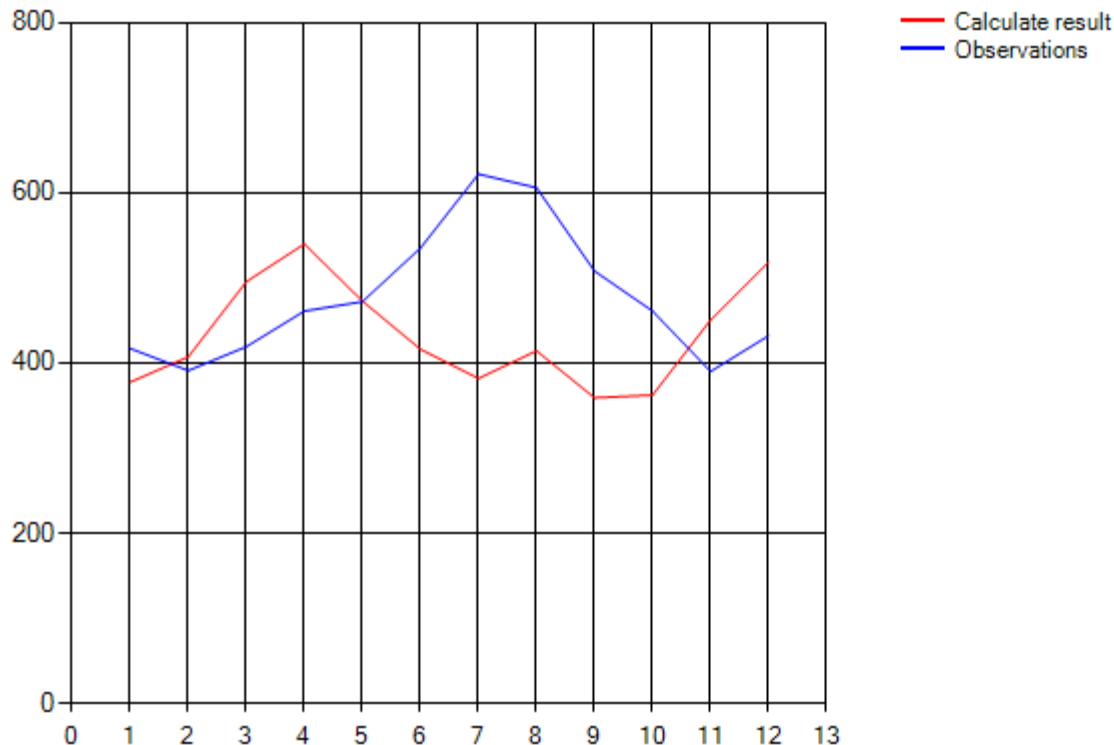
Áp dụng mạng neuron vào dự báo dữ liệu chuỗi thời gian

- Áp dụng các giải thuật huấn luyện mạng neuron để tìm ra một mạng nhận các giá trị y_t , y_{t-1} , y_{t-2} , ..., y_{t-s} làm đầu vào và sinh ra y_{t+1}

Áp dụng mạng neuron vào dự báo dữ liệu chuỗi thời gian



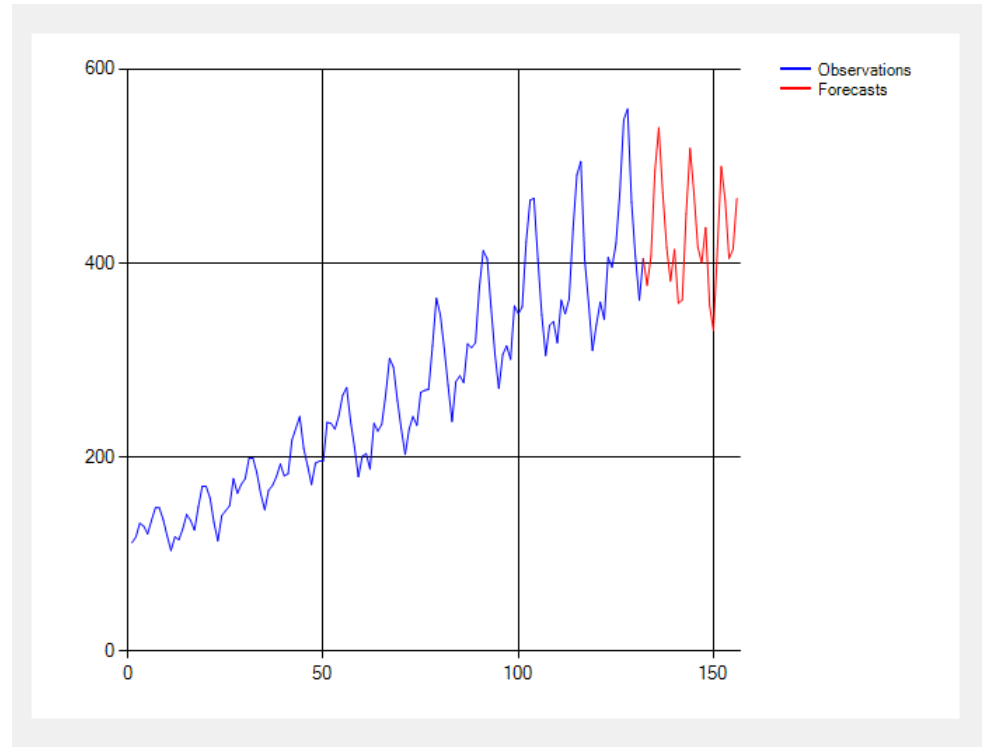
Áp dụng mạng neuron vào dự báo dữ liệu chuỗi thời gian



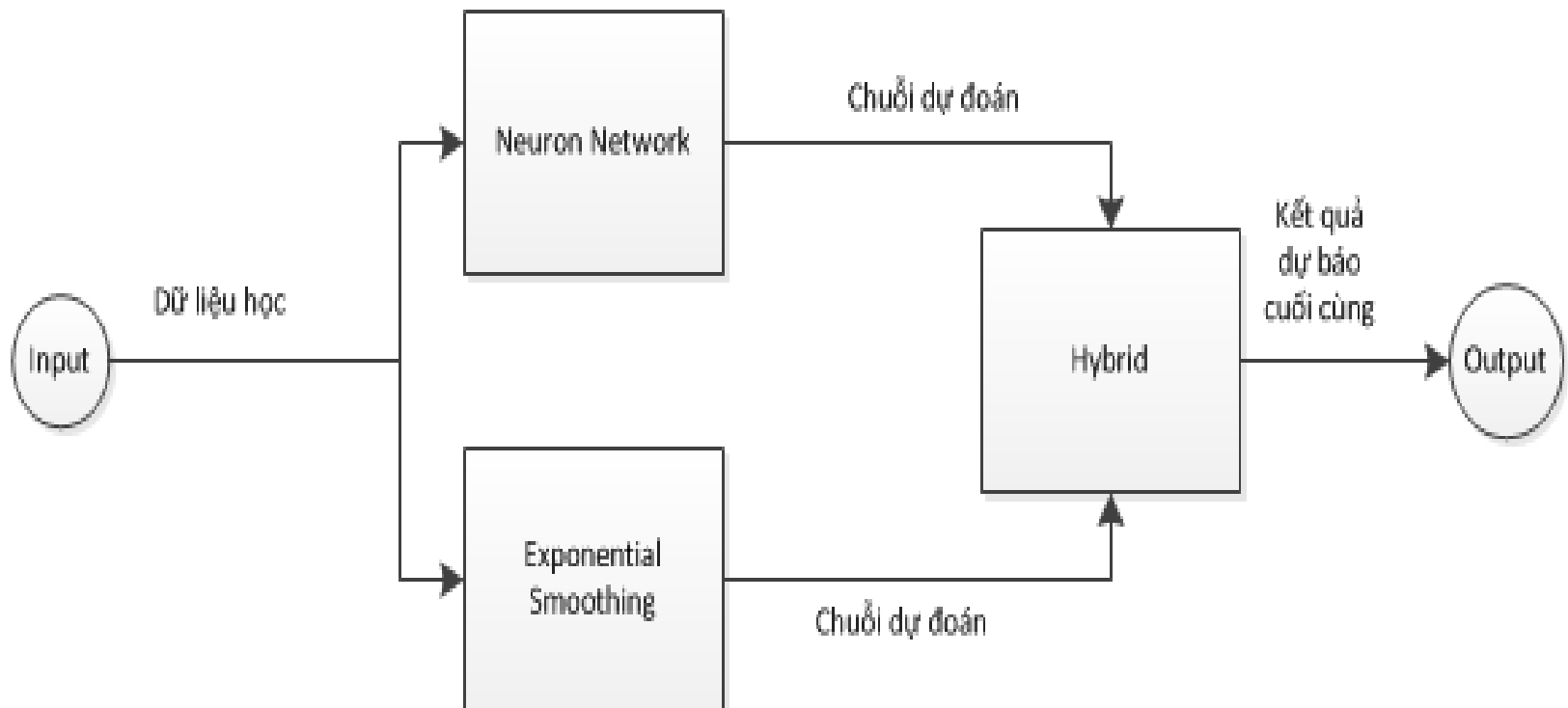
MAE: 96.4920339036461
SSE: 165745.771922352
MSE: 13812.147660196
MAPE: 19.0001199937667%

Áp dụng mạng neuron vào dự báo dữ liệu chuỗi thời gian

- Mạng neuron tuy có khả năng xấp xỉ tốt các hàm phi tuyến nhưng không thể mô hình tốt các chuỗi thời gian có tính xu hướng và tính mùa



Mô hình lai (Hybrid Model)



Môđun làm trơn lũy thừa (Exponential Smoothing)

$$\begin{aligned}Y_{n+k} &= (L_n + k \cdot T_n) \cdot S_{n+k-s} \\L_t &= \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \\T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \\S_t &= \gamma \frac{Y_t}{L_t} + (1 - \gamma)s_{t-s}\end{aligned}$$

Mô hình nhân

Môđun làm trơn lũy thừa (Exponential Smoothing)

$$\begin{aligned}Y_{n+k} &= (L_n + k \cdot T_n) + S_{n+k-s} \\L_t &= \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} \\&\quad + T_{t-1}) \\T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \\S_t &= \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}\end{aligned}$$

Mô hình cộng

Môđun làm trơn lũy thừa (Exponential Smoothing)

Ước lượng ba hệ số α , β , γ

- Vét cạn (Brute Force)
- Sử dụng giải thuật leo đồi (Hill Climbing)
 - ❑ Leo đồi dốc nhất (Steepest Ascent Hill Climbing)
 - ❑ Tôi luyện mô phỏng (Simulated Annealing)
- Sử dụng kết hợp hai phương pháp trên
- Sử dụng phần mềm R (thông qua phần mềm RAndFriend)

Môđun mạng Neuron nhân tạo (Neuron Network)

- Cấu trúc mạng:
 - ❑ Mạng Neuron truyền thẳng
 - ❑ Số node nhập bằng số node ẩn và bằng chu kì của chuỗi dữ liệu
- Giải thuật huấn luyện:
 - ❑ Giải thuật lan truyền ngược (Back Propagation)
 - ❑ Giải thuật RPROP (Resilient Propagation)

Môđun lai (Hybrid Module)

- Giá trị đầu vào của môđun lai là giá trị đầu ra của hai môđun: làm trơn lũy thừa và mạng neuron nhân tạo.

- Giá trị đầu ra được tính theo công thức

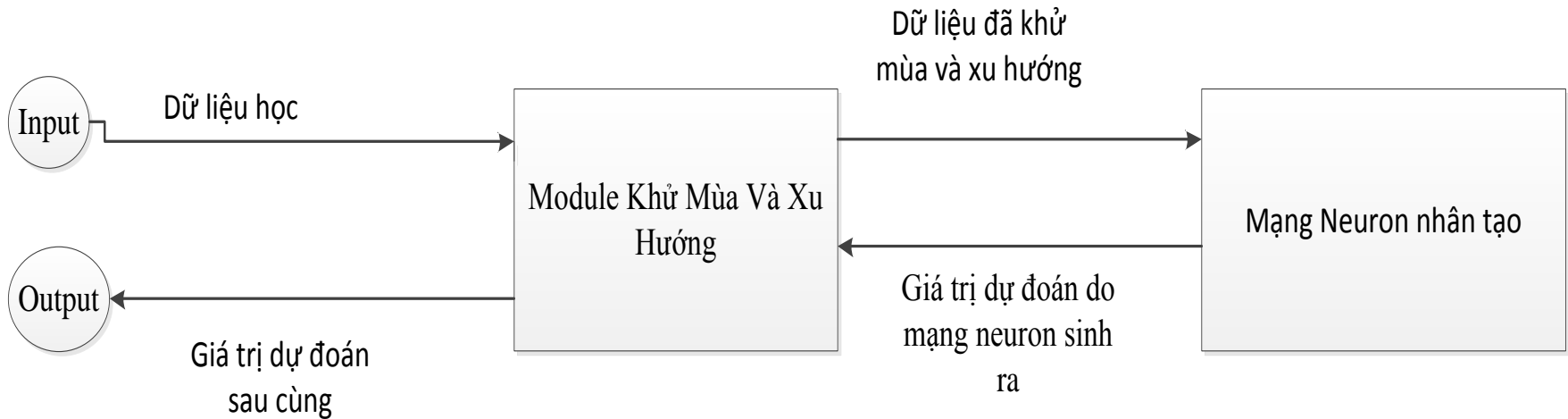
$$Y_{Hybrid} = \omega Y_{NN} + (1 - \omega) Y_{ES}$$

Trong đó: ω được gọi là trọng số lai

- Ước lượng giá trị của trọng số lai bằng cách tối thiểu hóa giá trị bình phương sai số lỗi:

$$MSE = \sum (Y_t - [\omega N_t + (1 - \omega) S_t])^2$$

Mô hình khử xu hướng, khử mùa



Môđun khử mùa và khử xu hướng

Hiện thực các kĩ thuật sau:

- Khử xu hướng

- ❑ Kĩ thuật khử xu hướng tuyến tính: ta xấp xỉ chuỗi thời gian bằng một đường thẳng hồi quy $at + b$ với t là biến thời gian. Ứng với mỗi t , lấy Y_t trừ đi $at+b$

- ❑ Kĩ thuật khử xu hướng bằng lấy hiệu: Với chuỗi thời gian $\{Y_t\}$ có tính xu hướng, đặt $X_t = Y_{t+1} - Y_t$ thì chuỗi thời gian $\{X_t\}$ sinh ra là một chuỗi không có tính xu hướng.

Môđun khử mùa và khử xu hướng

Hiện thực các kĩ thuật sau:

- Khử mùa

- Kĩ thuật khử mùa bằng lấy hiệu theo mùa:

- Kỹ thuật này thực hiện việc biến đổi chuỗi thời gian $\{Y_t\}$ thành chuỗi $\{X_t\}$ như sau $X_t = Y_{t+s} - Y_t$, với s là độ lớn một chu kỳ của chuỗi thời gian

Môđun khử mùa và khử xu hướng

Hiện thực các kĩ thuật sau:

- Khử mùa

- ❑ Kĩ thuật khử mùa bằng RTMA(ratio to moving average): Ta sẽ ước lượng *chỉ số mùa* (seasonal index) của các thời đoạn trong một chu kỳ của chuỗi thời gian rồi lấy giá trị của mỗi thời đoạn chia cho chỉ số mùa tương ứng của nó

Thực nghiệm

Chương trình thực nghiệm:

- Cả hai mô hình được hiện thực độc lập bằng ngôn ngữ lập trình C# trên nền .NET Framework 4.0
- Quá trình thực nghiệm được diễn ra trên máy có bộ vi xử lý Core 2 Duo, RAM 3GB, hệ điều hành Window 7 32bits được cài sẵn chương trình RAndFriends

Thực nghiệm

Dữ liệu thực nghiệm gồm:

Tên	Mô tả	Chu kì - Kích thước
Pan Am	Lượng khách hàng đặt chỗ hàng tháng của hãng hàng không Pan Am từ năm 1946 đến năm 1960	12-144
Mauna Loa	Mật độ khí cacbonic trong khí quyển hàng tháng ở Mauna Loa (Hawaii)	12-468
Lung Diseases	Số người chết trung bình hàng tháng vì bệnh phổi ở Anh	12-72
Queenslands	Doanh số bán hàng hàng tháng của một cửa hàng bán đồ lưu niệm ở Queensland, Australia, từ năm 1987 đến năm 1993	12-84
Gas	Lượng tiêu thụ khí đốt trung bình theo quý tại Anh	4-108

Thực nghiệm

Cách thức thực nghiệm: chạy các mô hình với sự thay đổi các thông số cấu hình.

❑ Mạng Neuron nhân tạo:

- Hai giải thuật RPROP – BP
- Số lượng tối đa epoches (1000-1500)

❑ Kỹ thuật làm trơn lũy thừa

- Ước lượng: sử dụng R, phương pháp kết hợp vét cạn và tối luyện mô phỏng
- Mô hình: mô hình cộng và mô hình nhân

Thực nghiệm

Cách thức thực nghiệm: chạy các mô hình với sự thay đổi các thông số cấu hình.

- ❑ Kỹ thuật khử xu hướng, khử mùa
 - Khử xu hướng: tuyến tính và lấy hiệu
 - Khử mùa: lấy hiệu và RTMA

Thực nghiệm

- Số lần chạy: mỗi cấu hình chạy ba lần và lấy kết quả trung bình, cấu hình cho kết quả dự đoán tốt nhất sẽ xem là kết quả của mô hình để so sánh với các mô hình khác
- Dữ liệu để đánh giá độ chính xác dự báo: chu kỳ cuối của chuỗi thời gian.
- Thông số đánh giá: MAPE, MAE

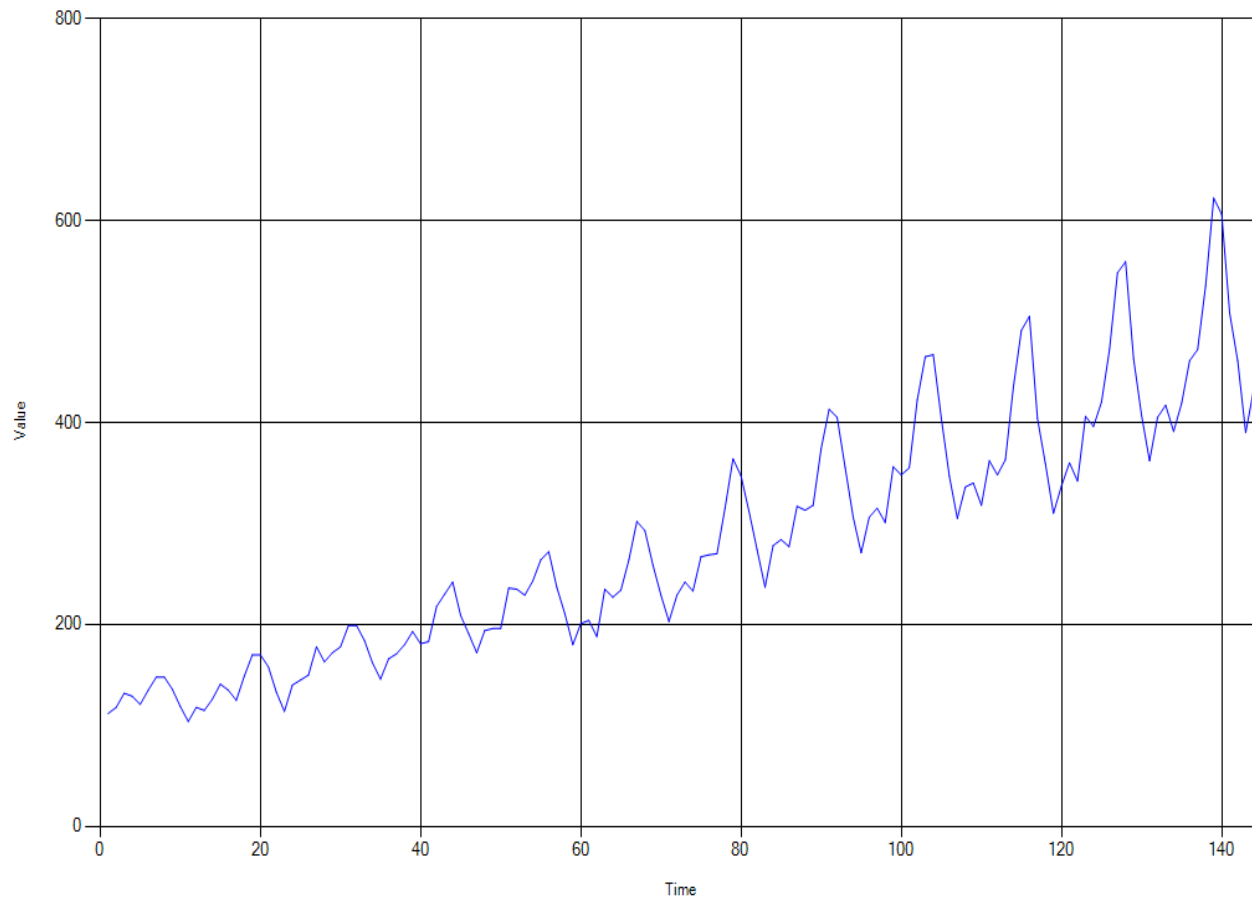
Thực nghiệm

- Sai số tuyệt đối trung bình *MAE* (mean absolute error) tính theo công thức:

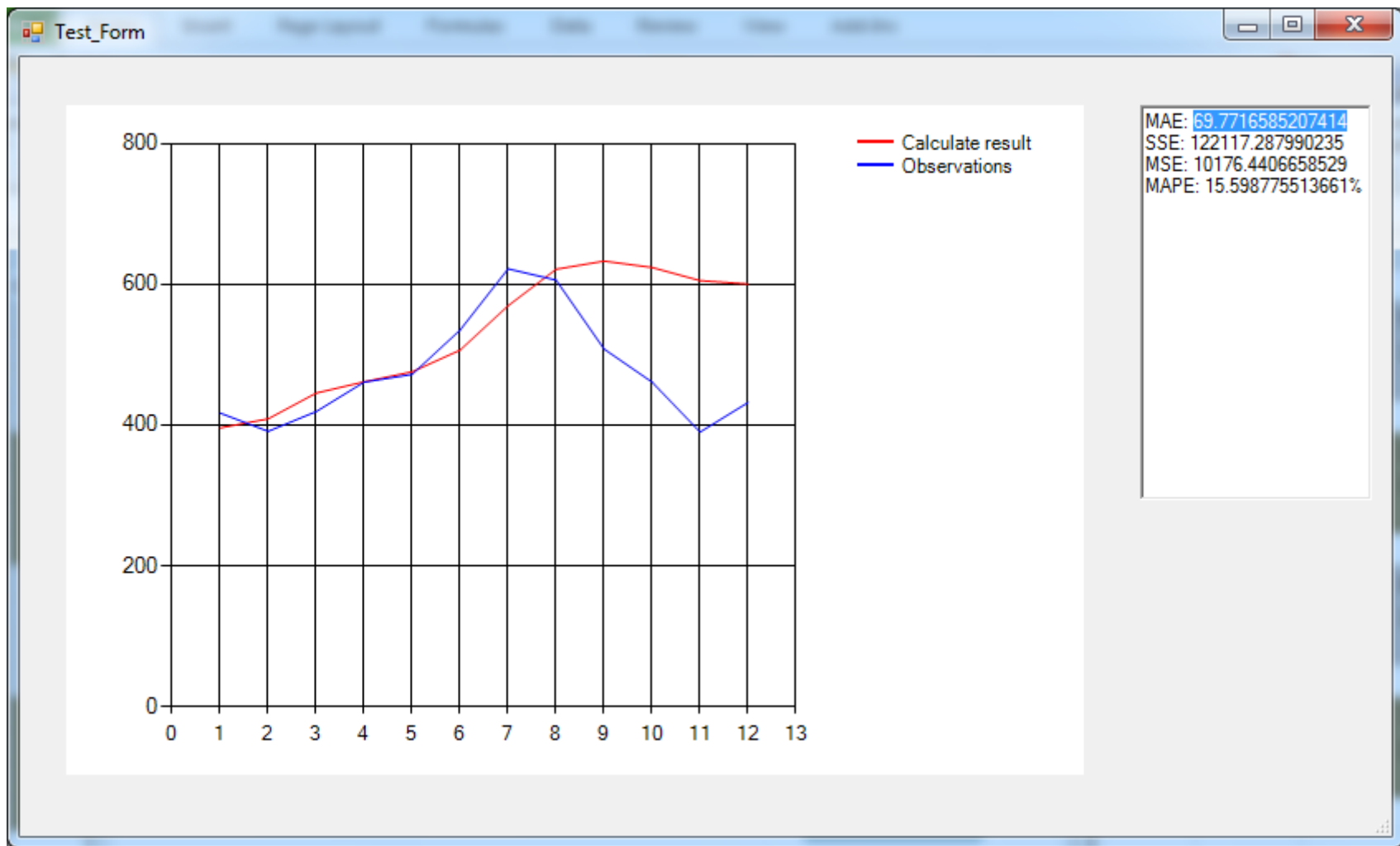
$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

- Sai số tuyệt đối phần trăm *MAPE* (mean absolute percentage error)

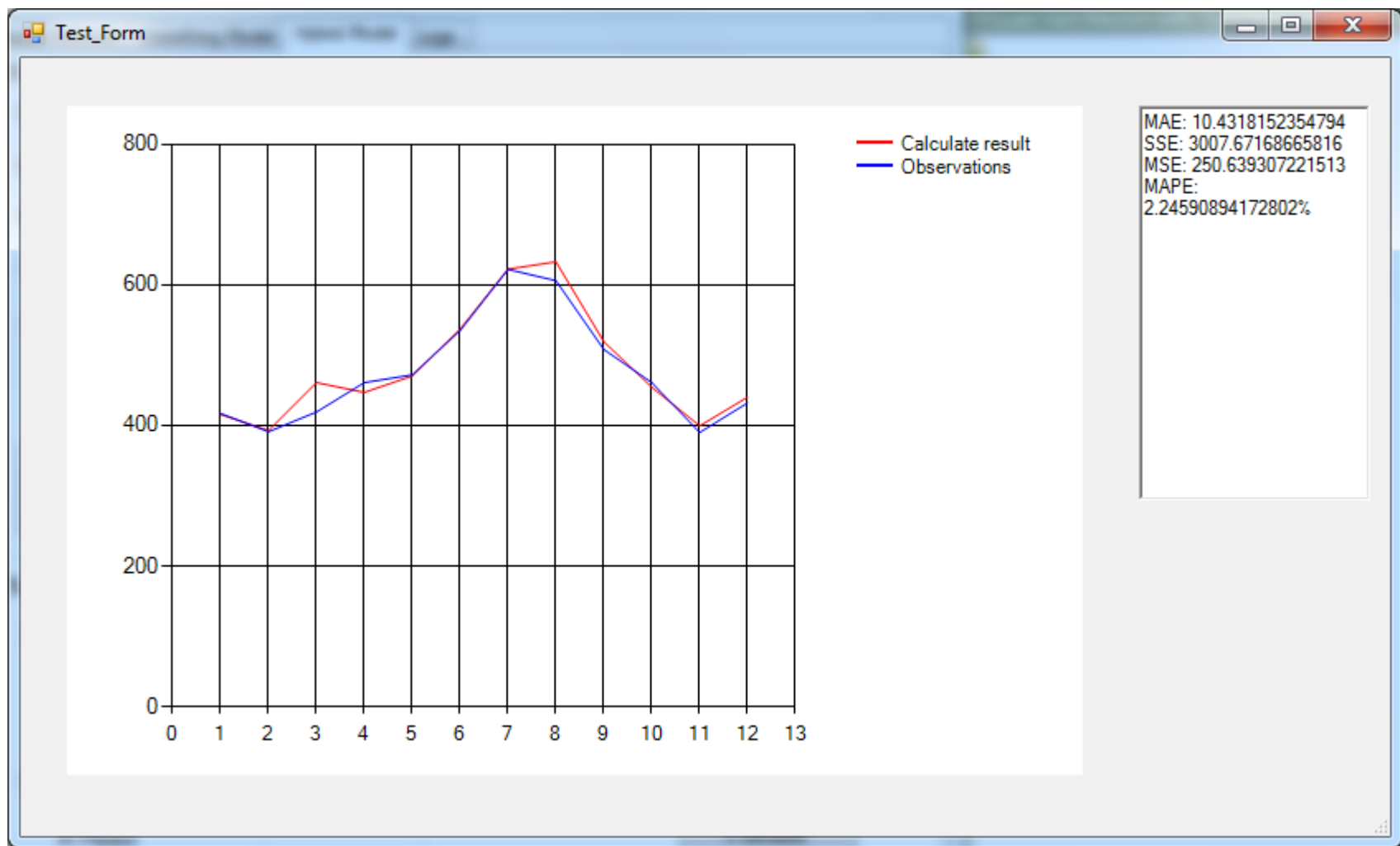
$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i}$$



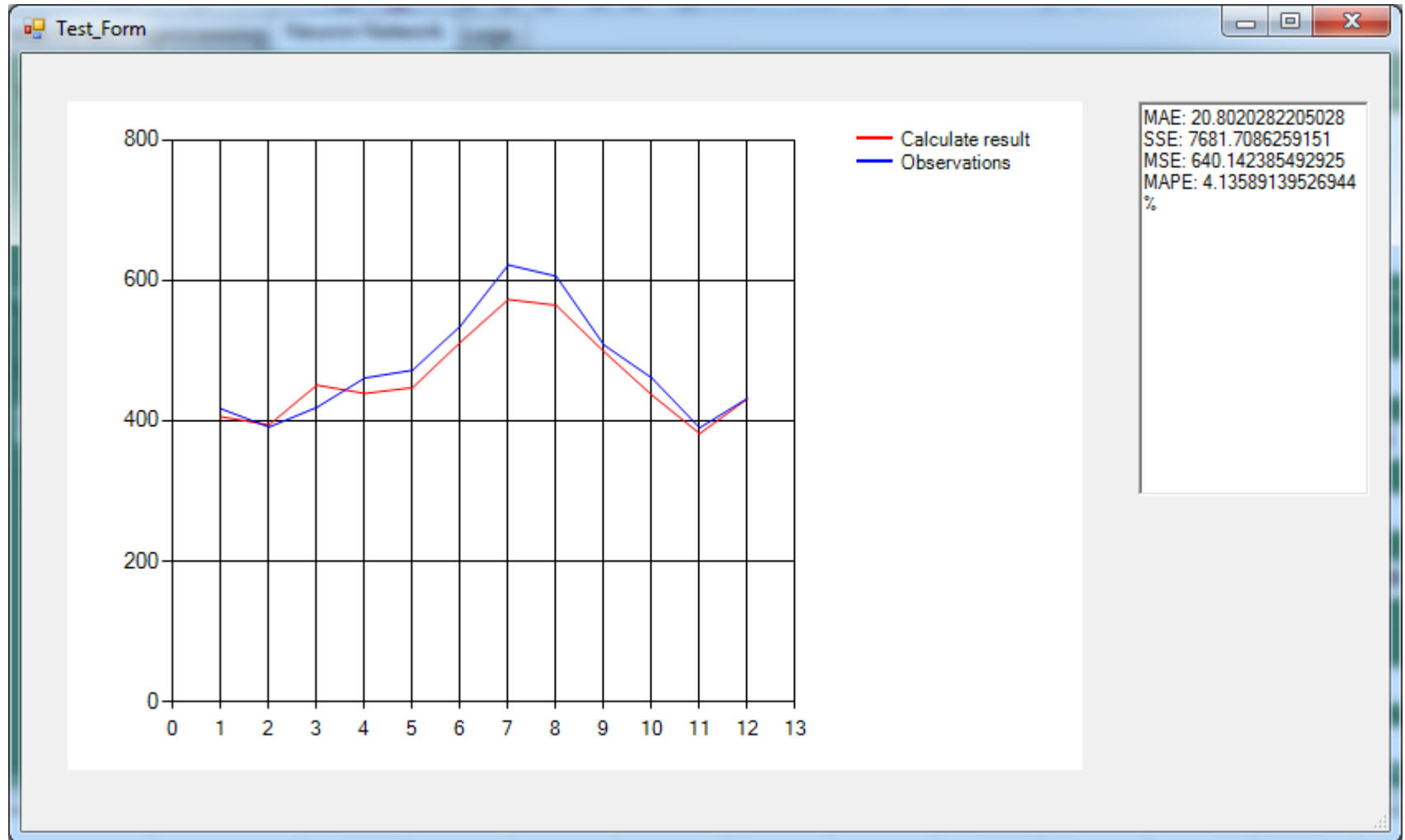
Đồ thị mô tả bộ dữ liệu Pan Am



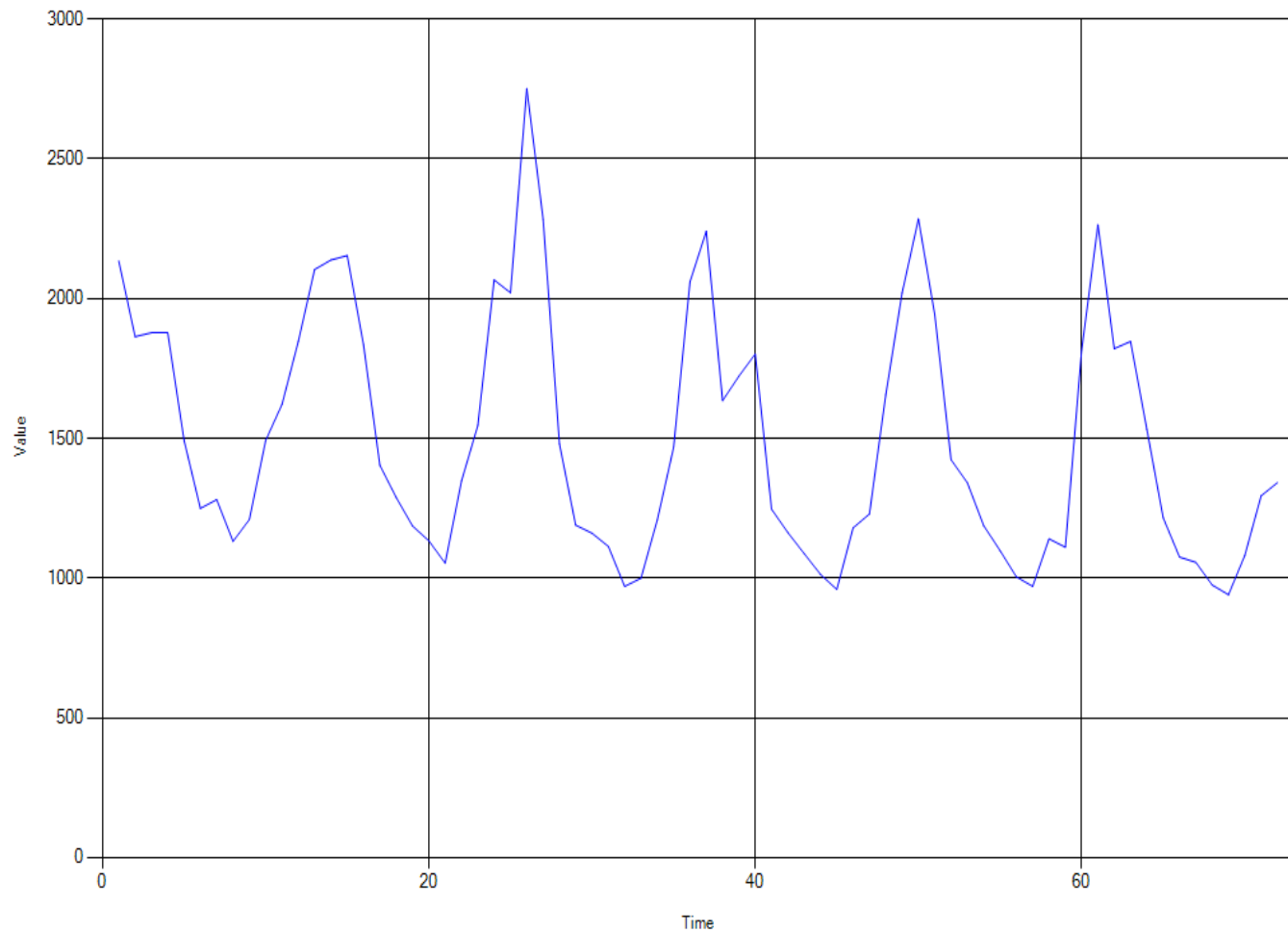
Kết quả dự báo chuỗi Pan Am bằng mạng Neuron
thuần



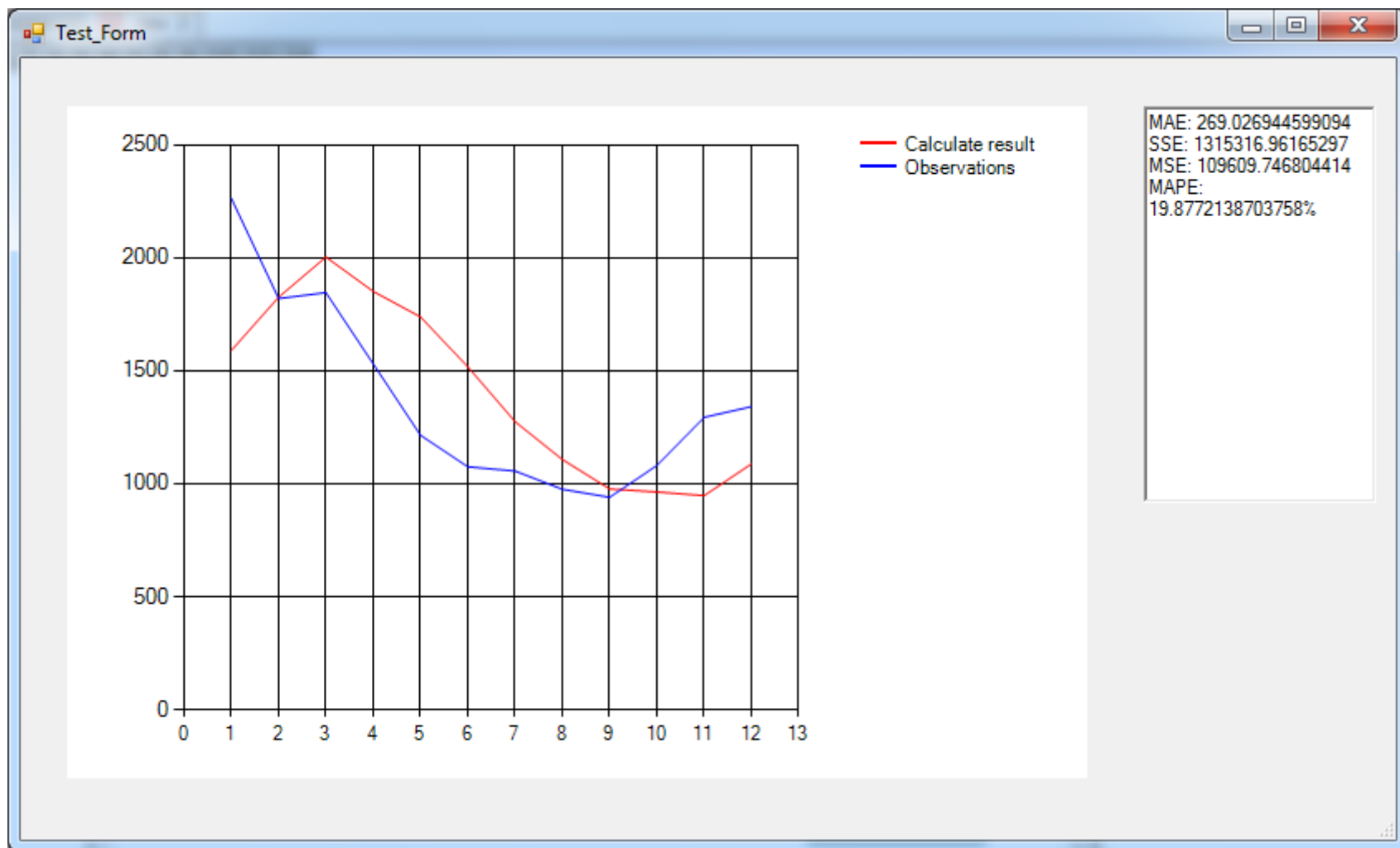
Kết quả dự báo chuỗi Pan Am bằng mô hình lai



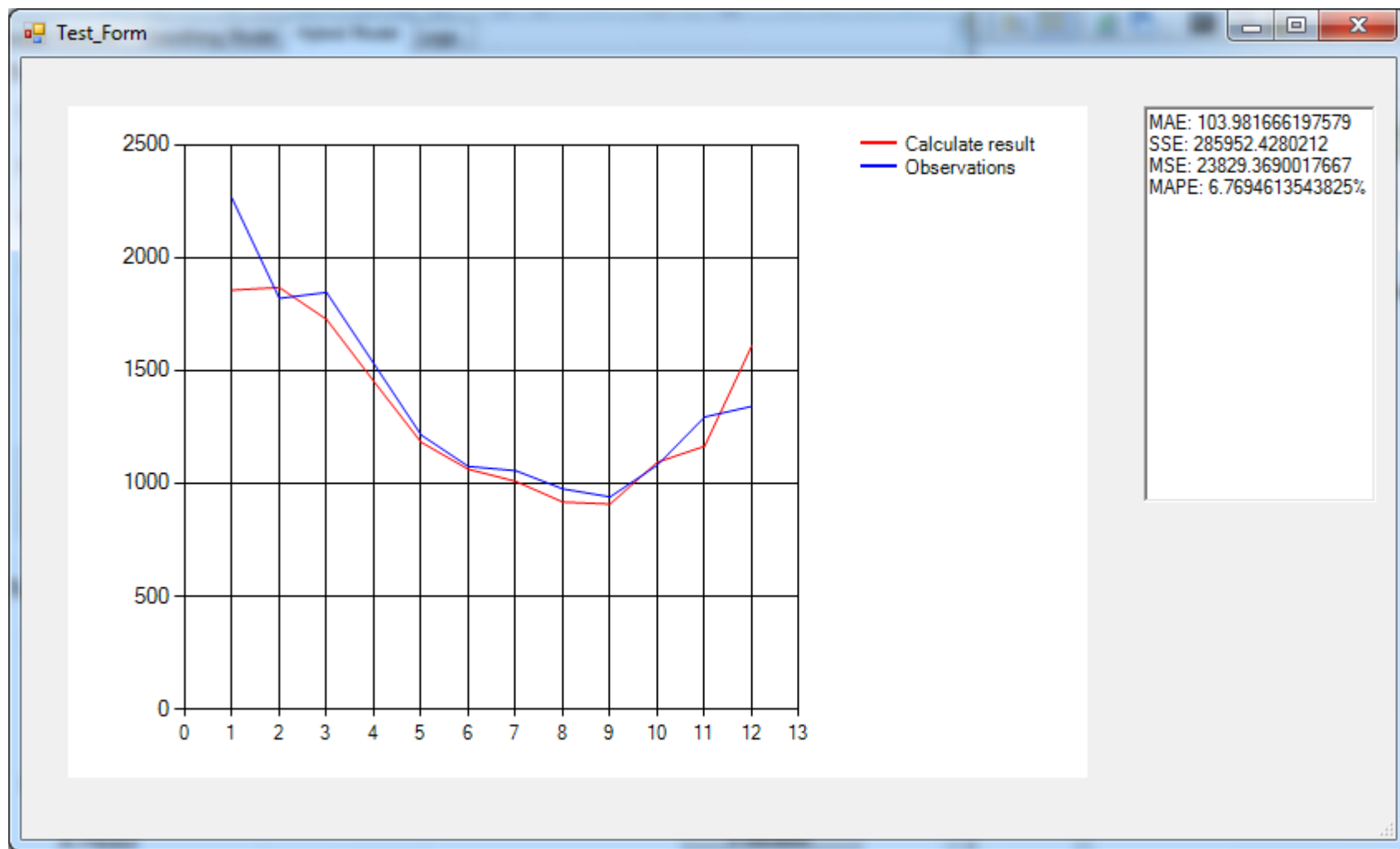
Kết quả dự báo chuỗi Pan Am bằng mô hình
khử mùa và xu hướng



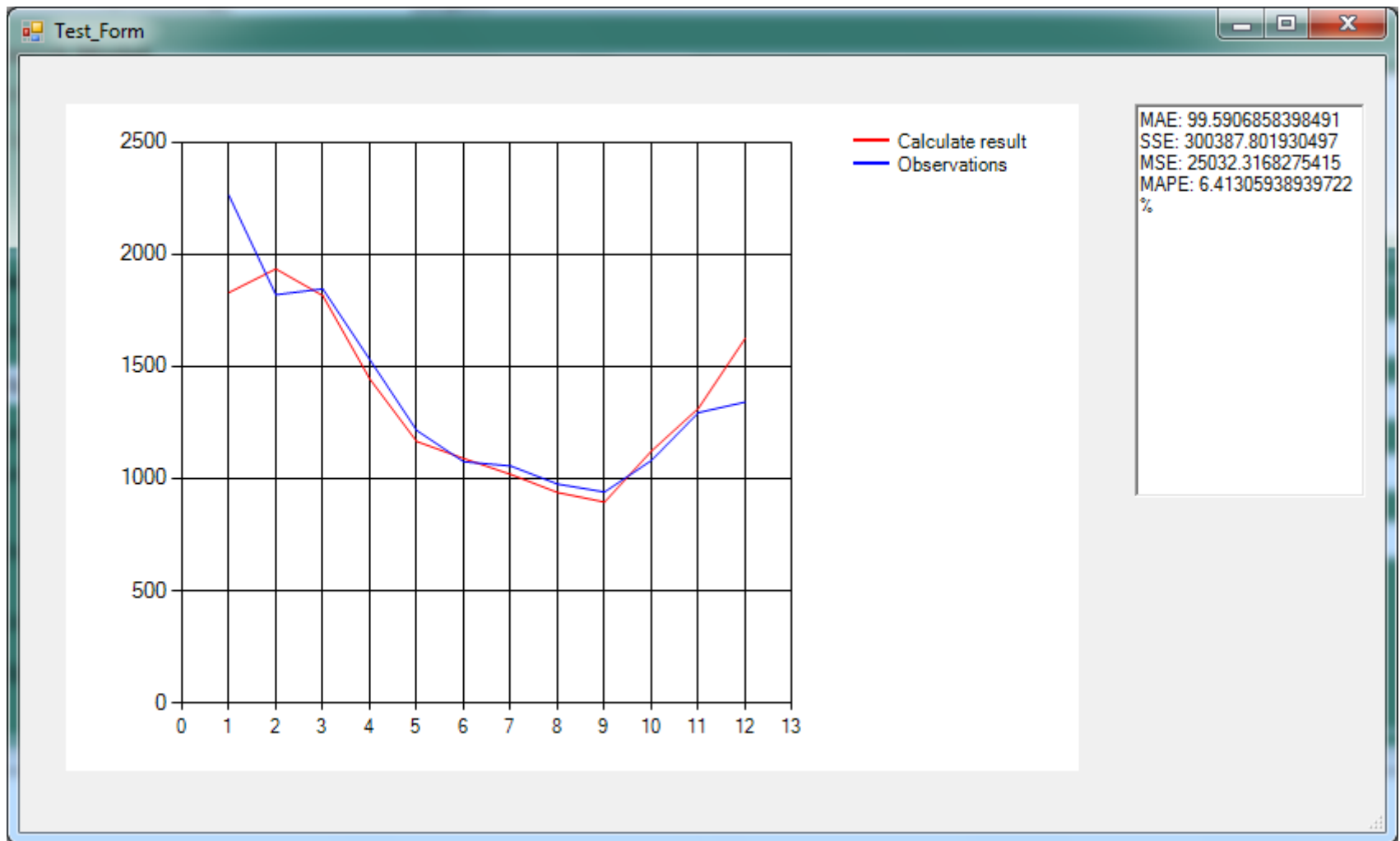
Đồ thị mô tả bộ dữ liệu Lung Diseases



Kết quả dự báo chuỗi Lung Diseases
bằng mạng Neuron thuần



Kết quả dự báo chuỗi Lung Diseases
bằng mô hình lai



Kết quả dự báo chuỗi Lung Diseases
bằng mô hình khử mùa và xu hướng

Bảng kết quả thực nghiệm

Tên dữ liệu	Neuron thuần	Mô hình Hybrid	Mô hình Preprocess
Pan Am	MAE: 72.38 MAPE: 16.23%	MAE: 10.43 MAPE: 2.25%	MAE: 23.90 MAPE: 4.69%
Mauna Loa	MAE: 2.37 MAPE: 0.65%	MAE: 0.34 MAPE: 0.09%	MAE: 0.33 MAPE: 0.09%
Lung Diseases	MAE: 274.66 MAPE: 20.37%	MAE: 103.98 MAPE: 6.77%	MAE: 98.40 MAPE: 6.41%
Queenslands	MAE: 20319.49 MAPE: 88.15%	MAE: 6584.78 MAPE: 19.68%	MAE: 3817.22 MAPE: 18.10%
Gas	MAE: 239.42 MAPE: 37.73%	MAE: 31.23 MAPE: 3.52%	MAE: 25.49 MAPE: 3.25%

Kết luận

Trong quá trình thực hiện đề tài, chúng tôi đã làm được những công việc sau:

- Áp dụng mạng neuron để dự báo đối với dữ liệu chuỗi thời gian.
- Hiện thực các phương pháp khử mùa và khử xu hướng đối với dữ liệu chuỗi thời gian như: khử xu hướng bằng phương pháp lấy hiệu, khử xu hướng bằng phương pháp tuyến tính, lấy hiệu theo mùa, khử mùa bằng phương pháp RTMA.

Kết luận

- Hiện thực kĩ thuật làm trơn lũy thừa như: làm trơn lũy thừa giản đơn, làm trơn lũy thừa Holt, làm trơn lũy thừa Winters. Ngoài ra, tiến hành nghiên cứu và tìm ra phương pháp ước lượng các hệ số trong kĩ thuật làm trơn lũy thừa Winters bằng việc kết hợp vét cạn và phương pháp tối luyện mô phỏng.
- Thực hiện kết hợp phần mềm R bằng R(D)COM trong chương trình C#.NET

Kết luận

- Nghiên cứu việc kết hợp hai kĩ thuật: khử mùa, khử xu hướng và làm trơn lũy thừa với mạng neuron nhằm nâng cao chất lượng dự báo đối với dữ liệu chuỗi thời gian có tính mùa và xu hướng.
- Tiến hành hiện thực hai mô hình dự báo từ các nghiên cứu trên.

Kết luận

- Tiến hành chạy thực nghiệm với năm bộ dữ liệu thực tế và đánh giá, kiểm chứng tính đúng đắn của cơ sở lý thuyết cũng như quá trình hiện thực. Kết quả hai mô hình đề xuất cho kết quả dự báo tốt hơn mạng neuron nhân tạo cho chuỗi thời gian có tính mùa và xu hướng

Hướng phát triển

- Đối với mô hình lai, thay thế phương thức khởi tạo các thông số đối với mô hình làm trơn lũy thừa đang dùng bằng các phương pháp tiên tiến hơn như phương pháp *dựa trên hồi quy* (regression-based procedure) hay *phương pháp dựa trên phân giải* (decomposition-based) để có thể đưa ra dự báo chính xác hơn.

Hướng phát triển

- Đối với mô hình khử mùa, khử xu hướng kết hợp mạng neuron, cải tiến phương pháp khử xu hướng tuyến tính để có thể áp dụng tốt cho các chuỗi thời gian có xu hướng mang hình dạng đường cong. Áp dụng các kỹ thuật khử mùa tiên tiến gần đây như kỹ thuật X-12-ARIMA vào chương trình

Tài liệu tham khảo

- G. Zhang, M. Qi. *Trend Time-Series Modeling And Forecasting With Neural Networks*. IEEE Transactions on Neural Network, vol. 19, no. 5, pages 808-816, 2008.
- G. Zhang, D. M. Kline. *Quarterly Time-Series Forecasting With Neural Networks*. IEEE Transactions on Neural Network, vol. 18, no. 6, pages 1800-1814, 2007.
- K. Lai, L. Yu, S. Wang, W. Huang. *Hybridizing Exponential Smoothing And Neural Network For Financial Time Series Predication*. ICCS'06 Proceedings of the 6th international conference on Computational Science, vol. 4, pages 493-500, 2006.
- G. Zhang, M. Qi. *Neural Network Forecasting For Seasonal And Trend Time Series*. European Journal of Operational Research vol. 160, pages 501-514, 2005.
- J. E. Hanke, D. W. Wichern. *Business Forecasting*, Pearson Prentice Hall, ISBN 0-13-141290-6, 2005.
- Trần Đức Minh. Luận văn thạc sĩ *Mạng Neural Truyền Thẳng Và Ứng Dụng Trong Dự Báo Dữ Liệu*. Đại học quốc gia Hà Nội, 2002
- F. Virili, B. Freisleben. *Preprocessing Seasonal Time Series For Improving Neural Network Predictions*. Proceesings of CIMA 99 Computational Intelligence Methods and Applications, Rochester-NY, pages 622-628, 1999.
- G. Zhang, M. Y. Hu. *Neural Network Forecasting Of The British Pound/US Dollar Exchange Rate*. Omega, International Journal of Management Science, 26, pages 495-506, 1998.
- T. M. Mitchell. *Machine Learning*, McGraw-Hill Science/ Engineering/ Math, ISBN 0070428077, 1997.
- I. Kaastra, M. Boyd. *Designing A Neural Network For Forecasting Financial And Economic Time Series*. Neurocomputing, vol. 10, pages 215-236, 1996.
- M. Riedmiller. *Advanced Supervised Learning In Multi-layer Perceptrons – From Backpropagation To Adaptive Learning Algorithms*. Int. Journal of Computer Standards and Interfaces, 1994.

Q & A

Thank you!