Nội dung

[CÁC CẢI TIẾN ĐÃ THỰC HIỆN 3](#_Toc285058276)

[BỔ SUNG VẾ ĐỘ ĐO 3](#_Toc285058277)

[THỰC NGHIỆM 3](#_Toc285058278)

[Mô tả dữ liệu 3](#_Toc285058279)

[Kết quả 3](#_Toc285058280)

[KẾT LUẬN 7](#_Toc285058281)

[CÔNG VIỆC TƯƠNG LAI 8](#_Toc285058282)

[TÀI LIỆU THAM KHẢO 8](#_Toc285058283)

Báo cáo kết quả dự đoán xu hướng

# CÁC CẢI TIẾN ĐÃ THỰC HIỆN

Thêm chỉ số kỹ thuật vào đầu vào bao gồm: Trung bình trượt giản đơn, Trung bình trượt số mũ, MACD.

Thực hiện huấn luyện multi-period (test đến đâu, huấn luyện đến đó).

Dùng GA cho việc chọn tham số của SVR.

# BỔ SUNG VẾ ĐỘ ĐO

Do vấn đề về giá đóng cửa có thể không thay đổi trong nhiều phiên giao dịch liên tiếp nên nhóm sử dụng thêm độ đo sau để có cái nhìn rõ hơn về kết quả dự đoán. Độ đo do nhóm dự đoán giá phát biểu:

DM4Price (Direction Measure For Price):

Xét tích

* + - Nếu tích > 0: dự đoán đúng.
    - Nếu tích < 0: dự đoán sai.
    - Nếu tích = 0: không đánh giá được.

Độ đo DM4Price sẽ cho ra 3 kết quả lần lượt ứng với phần trăm số phần tử dự đoán đúng, dự đoán sai và không đánh giá được.

# THỰC NGHIỆM

## Mô tả dữ liệu

* Nước ngoài:
  + NYSE: từ 9/1/2004 đến 31/12/2007; gồm 1001 điểm dữ liệu.[4]
* Việt Nam:
  + FTP: từ 13/12/2006 đến 9/8/2010; gồm 914 điểm dữ liệu.
  + DHG: từ 21/12/2006 đến 10/12/2010; gồm 994 điểm dữ liệu.
  + VIS: từ 25/12/2006 đến 10/12/2010; gồm 988 điểm dữ liệu.
  + VNM: từ 19/1/2006 đến 10/12/2010; gồm 1222 điểm dữ liệu.
  + BT6: từ 18/04/2002 đến 10/12/2010; gồm 2164 điểm dữ liệu.

## Kết quả

Kết quả kiểm thử đối với mã NYSE trên period = 1

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | NMSE | Sign(%) |
| SVR | Grid | 96.74 | 62 |
| SVR cải tiến | GA | 98.2 | 63.6 |
| Multi-period | 100.2 | 58.1 |
| Grid search | 97.74 | 63.55 |
| ANN |  | 103 | 54 |
| Mô hình [1] |  | 99.41 | 68.69 |

Nhận xét: Multi-period cho SVR tỏ ra không hiệu quả, trong khi GA cho kết quả ngang với Grid search trong việc tìm tham số. Phần cải tiến có nhỉnh hơn phần SVR ban đầu hơn một điểm phần trăm.

Kết quả trên các mã Việt Nam cũng cho thấy một bức tranh tương tự. Do vậy để ngắn gọn, báo cáo không trình bày phần multi-period ở đây.

* Period = 1:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ANN | | | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | | Độ lỗi khi train | |
| Right | Wrong | Cannot measure | NMSE | Max Loop |
| BT6 | 1.28 | 42 | 72.09 | 19.06 | 8.83 | 1.65 | 2000 |
| DHG | 1.34 | 25 | 51.02 | 13.27 | 35.71 | 1.03 | 2000 |
| FPT | 1.3 | 43 | 72.73 | 23.23 | 4.04 | 1.18 | 2000 |
| VIS | 1.03 | 41 | 65.31 | 34.69 | 0 | 1.04 | 2000 |
| VNM | 1.15 | 32 | 57.85 | 21.48 | 20.66 | 0.98 | 2000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR – Grid search | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.28 | 41 | 72.62 | 18.55 | 8.83 |
| DHG | 1.27 | 40 | 56.5 | 7.7 | 35.71 |
| FPT | 1.02 | 45 | 75.87 | 20.09 | 4.04 |
| VIS | 1.07 | 55 | 71.94 | 28.06 | 0 |
| VNM | 1.01 | 33 | 69.13 | 10.2 | 20.66 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR – GA | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.24 | 44 | 72.22 | 18.95 | 8.83 |
| DHG | 1.13 | 40 | 54.5 | 9.7 | 35.71 |
| FPT | 1.34 | 44 | 75.87 | 20.09 | 4.04 |
| VIS | 1.02 | 54 | 70.9 | 29.1 | 0 |
| VNM | 1.12 | 38 | 70.2 | 9.12 | 20.66 |

* Period = 5:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ANN | | | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | | Độ lỗi khi train | |
| Right | Wrong | Cannot measure | NMSE | Max Loop |
| BT6 | 1.4 | 44 | 58.69 | 32.39 | 8.92 | 1.38 | 2000 |
| DHG | 1.17 | 25 | 41.67 | 27.08 | 31.25 | 1.01 | 2000 |
| FPT | 1.21 | 38 | 56.7 | 42.27 | 1.03 | 1.78 | 2000 |
| VIS | 1.03 | 70 | 62.11 | 37.89 | 0 | 1 | 2000 |
| VNM | 1.78 | 50 | 48.74 | 30.25 | 21 | 1.04 | 2000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR – Grid search | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.29 | 45 | 58.92 | 32.16 | 8.92 |
| DHG | 1.14 | 31 | 55.1 | 13.65 | 31.25 |
| FPT | 1.04 | 42 | 52.97 | 45.9 | 1.03 |
| VIS | 1.8 | 51 | 54.97 | 45.02 | 0 |
| VNM | 1.21 | 45 | 55.71 | 23.27 | 21 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR– GA | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.15 | 44 | 58.1 | 32.98 | 8.92 |
| DHG | 1.21 | 32 | 55.35 | 13.4 | 31.25 |
| FPT | 1.24 | 41 | 52.23 | 46.64 | 1.03 |
| VIS | 1.14 | 50 | 55 | 45. | 0 |
| VNM | 1.53 | 45 | 54 | 25 | 21 |

* Period = 10:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ANN | | | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | | Độ lỗi khi train | |
| Right | Wrong | Cannot measure | NMSE | Max Loop |
| BT6 | 2.32 | 39 | 51.42 | 42.86 | 5.71 | 1.03 | 2000 |
| DHG | 1.09 | 25 | 38.7 | 29.03 | 32.25 | 0.98 | 2000 |
| FPT | 1.41 | 61 | 58.51 | 36.17 | 5.31 | 2.18 | 2000 |
| VIS | 0.93 | 73 | 50 | 50 | 0 | 1.07 | 2000 |
| VNM | 1.35 | 25 | 38.79 | 46.55 | 14.66 | 1.12 | 2000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR – Grid search | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.67 | 44 | 56.66 | 37.62 | 5.71 |
| DHG | 1.4 | 31 | 48.45 | 19.29 | 32.25 |
| FPT | 1.18 | 35 | 52.04 | 42.64 | 5.31 |
| VIS | 1.66 | 46 | 52.43 | 47.56 | 0 |
| VNM | 0.99 | 52 | 48.33 | 37 | 14.66 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR– GA | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.26 | 45 | 55.12 | 39.16 | 5.71 |
| DHG | 1.56 | 28 | 46 | 21.74 | 32.25 |
| FPT | 1.32 | 40 | 52.2 | 42.48 | 5.31 |
| VIS | 1.59 | 57 | 52.13 | 47.86 | 0 |
| VNM | 1.03 | 51 | 45.3 | 40.03 | 14.66 |

* Period = 30:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ANN | | | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | | Độ lỗi khi train | |
| Right | Wrong | Cannot measure | NMSE | Max Loop |
| BT6 | 0.91 | 46 | 51.01 | 42.93 | 6.06 | 0.98 | 2000 |
| DHG | 2.57 | 68 | 39.5 | 29.63 | 30.86 | 1.15 | 2000 |
| FPT | 1.89 | 55 | 57.31 | 40.24 | 2.43 | 1.04 | 2000 |
| VIS | 1.18 | 79 | 48.75 | 51.25 | 0 | 1.04 | 2000 |
| VNM | 1.41 | 52 | 44.23 | 35.58 | 20.19 | 1.17 | 2000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR – Grid search | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.84 | 47 | 50.75 | 43.18 | 6.06 |
| DHG | 1.89 | 48 | 45.16 | 23.98 | 30.86 |
| FPT | 3.47 | 56 | 54.78 | 42.78 | 2.43 |
| VIS | 1.6 | 41 | 49.03 | 50.93 | 0 |
| VNM | 3.6 | 52 | 45.26 | 34.54 | 20.19 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SVR– GA | | | | | |
|  | NMSE | SIGN(%) | DM4Price(%) | | |
| Right | Wrong | Cannot measure |
| BT6 | 1.62 | 48 | 50 | 43.94 | 6.06 |
| DHG | 1.54 | 46 | 44.9 | 24.25 | 30.86 |
| FPT | 2.12 | 55 | 54.59 | 42.98 | 2.43 |
| VIS | 1.9 | 40 | 46.2 | 53.78 | 0 |
| VNM | 3.76 | 52 | 48.2 | 31.61 | 20.19 |

Ta thấy, mặc dù có cải tiến, nhưng kết quả đạt được vấn không cao. SVR vẫn nhỉnh hơn ANN tuy nhiên kết quả chỉ khiêm tốn xoay quanh mức 50%.

# KẾT LUẬN

Về phần SVR cho xu hướng, sự so sánh tương quan kết quả trong [1] và kiểm thử với nhiều cổ phiếu Việt Nam, ta rút ra được một vài kết luận sau đây:

* So sánh với [1], kết quả cải tiến cho một cải thiện nho nhỏ về độ chính xác. GA và Grid search chưa cho thấy sự vượt trội hẳn về kết quả dự đoán.
* Với các mã Việt Nam, phần đầu vào có cải thiện kết quả chính xác. Tuy nhiên, phần trăm chính xác vẫn chưa thỏa mãn, GA và Grid vẫn ở thế giằng co.

Về phần ANN, kết quả vẫn kém hơn SVR dù có một số cải thiện trong tính toán. Bên cạnh đó, thỉnh thoảng kết quả ANN lại tốt hơn ở period cao.

Có thể nói, sự không cải thiện kết quả ở các cổ phiếu Việt Nam phần nhiều là do sự non trẻ của thị trường. Các phiên giao dịch có biểu đồ giá nằm ngang dẫn đến việc dự đoán sai. Hơn nữa, trong mô hình hồi quy, đặc biệt là SVR, không thể định được ngưỡng để xác định khi nào giá không đổi. Nguyên nhân là do tính tổng quát hóa ở mô hình này nên đường hồi quy chỉ giao động vây quanh 0.

# CÔNG VIỆC TƯƠNG LAI

Từ kết luận trên, ta rút ra được một vài điều cần làm sắp tới như sau:

* Thay đổi để mô hình nhận hẳn kết quả đầu ra là -1 0 1.
* Để có được kết quả đầu ra như trên, ANN với 1 node đầu ra sẽ chuyển thành 3 node, SVR sẽ không được sử dụng mà thay bằng SVC (SVM cho phân lớp).
* Tuy nhiên vẫn cần có một định nghĩa chặt chẽ cho xu hướng để đảm bảo bài toán giải quyết tốt vấn đề chứng khoán thực tế ( Xu hương tăng hay giảm không hẳn là giá tăng hay giảm sau một chu kỳ). Trong [2], xu hướng tăng (giảm) được định nghĩa chặt chẽ như sau:
  + Giá đóng cửa phải cao hơn (thấp hơn) giá trung bình trượt 25 ngày.
  + Giá trung bình trượt 25 ngày phải cao hơn (thấp hơn) giá trung bình trượt 65 ngày.
  + Giá trung bình 25 ngày phải tăng (giảm) trong ít nhất 5 ngày.
  + Giá trung bình 65 ngày phải tăng (giảm) trong ít nhất 1 ngày.

Xu hương tăng (giảm) phải thỏa 4 tiêu chí trên, nếu không giá sẽ ở trạng thái không có xu hướng.

* Sau khi có được mô hình trên, sẽ lần lượt tiến hành them vào các cải tiến đã đề cập, để đảm bảo mô hình hoàn thiện.

# TÀI LIỆU THAM KHẢO

[1] ShiyiChen, Kiho Jeong,Wolfgang K.Hardle,*Recurrent Support Vector Regression for a Nonlinear ARMA Model with Applications to Forecasting Financial Returns*, SFB 649 Discussion Paper, 2008

[2] Binoy B. Nair, V.P Mohandas, N.R. Sakthivel, *A Genetic Algorithm Optimized Decision Tree-SVM based Stock Market Trend Prediction System,* IJCSE Vol.02, No. 09, 2010.