TỔNG HỢP VÀI VẤN ĐỂ VỀ MÔ HÌNH

Về độ đo

Có 3 độ đo thường được sử dụng trong các báo cáo về SVM: Mean squared error(MSE), Normalized mean squared error(NSME), Squared correlation coefficient(SCC). Trong đó MSE thể hiện sự sai khác cục bộ còn SCC thể hiện sự sai khác toàn cục.

Trong đó là giá trị đầu ra thực sự, là giá trị đầu ra dự đoán, , là các giá trị trung bình của chúng.

Ngoài ra, để tính hiệu quả dự đoán xu hướng của mô hình , người ta thường sử dụng công thức tính độ chính xác sau:

Về dữ liệu đầu vào

Khá nhiều mô hình đề xuất một số loại dữ liệu đầu vào khác ngoài giá đóng cửa. Tuy nhiên, chưa có công trình nào thực sự nghiên cứu sâu về vấn đề này để có thể trả lời câu hỏi đâu là những đầu vào tối ưu nhất. Hơn nữa, với những thị trường khác nhau, câu trả lời cho vấn đề trên không bất biến. Vì vậy, tại thời điểm này, đầu vào dành cho mô hình sẽ chỉ là giá đóng cửa. Mọi cải tiến về đầu vào sẽ được tìm hiểu sâu hơn trong giai đoạn sau.

Về việc chọn dữ liệu, ngoài các mã trong nước, cần có dữ liệu của thị trường nước ngoài mà các bài báo đã từng áp dụng. Ở đây đề xuất sử dụng thêm các mã chứng khoán của thị trường Đài Loan.

Đối với mô hình dự đoán xu hướng, giá đóng cửa phải được tinh chỉnh về dạng return. Đây sẽ là dữ liệu đầu vào thực tế:

 với I là giá

Về mô hình

Áp dụng ε-SVR với hàm Gaussian kernel và hàm tổn thất là ε-insensitive loss function.

Về các tham số

Đối với mô hình ε-SVR như trên, có 3 tham số quan trọng cần được xác định là C (regularization parameter), ε, và σ2. Để xác định các tham số này, phương pháp cơ bản nhất là dùng cross validation.

Về việc kiểm thử

Sử dụng 2 kiểu kiểm thử:

* + - Two-period: dữ liệu chia 2 phần, 1 phần huấn luyện, 1 phần kiểm (giống các nghiên cứu khác).
    - Multi-period: dữ liệu huấn luyện là phần dữ liệu liền trước của phần muốn kiểm thử.

Chẳng hạn có 100 ngày, để dự đoán ngày thứ 15 ta dùng dữ liệu từ ngày 1-14. Để dự đoán ngày 16 ta dùng dữ liệu từ ngày 2-15, vân vân…

Trong mỗi kiểu kiểm thử sẽ chia ra theo nhiều chu kỳ khác nhau: 5, 10, 30, 60 ngày.