

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM**



NGUYỄN THỊ THU HIỀN

**DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN
DỰA TRÊN HỆ SUY LUẬN ANFIS**

LUẬN VĂN THẠC SĨ

Chuyên ngành : Công nghệ thông tin

Mã số ngành: 60480201

TP. HỒ CHÍ MINH, tháng 01 năm 2016

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM**



NGUYỄN THỊ THU HIỀN

**DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN
DỰA TRÊN HỆ SUY LUẬN ANFIS**

LUẬN VĂN THẠC SĨ

Chuyên ngành : Công nghệ thông tin

Mã số ngành: 60480201

CÁN BỘ HƯỚNG DẪN KHOA HỌC: PGS.TS. QUẢN THÀNH THƠ

TP. HỒ CHÍ MINH, tháng 01 năm 2016

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH TẠI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM**

Cán bộ hướng dẫn khoa học : PGS. TS. Quản Thành Thor

Luận văn Thạc sĩ được bảo vệ tại Trường Đại học Công nghệ TP. HCM
ngày 10 tháng 09 năm 2016.

Thành phần Hội đồng đánh giá Luận văn Thạc sĩ gồm:

TT	Họ và tên	Chức danh Hội đồng
1	PGS.TS Võ Đình Bảy	Chủ tịch
2	TS. Lê Văn Quốc Anh	Phản biện 1
3	TS. Nguyễn Thị Thúy Loan	Phản biện 2
4	TS. Trần Đức Khánh	Ủy viên
5	TS. Cao Tùng Anh	Ủy viên, Thư ký

Xác nhận của Chủ tịch Hội đồng đánh giá Luận sau khi Luận văn đã được
sửa chữa (nếu có).

Chủ tịch Hội đồng đánh giá LV

TRƯỜNG ĐH CÔNG NGHỆ TP. HCM
PHÒNG QLKH – ĐTSĐH

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

TP. HCM, ngày 23 tháng 01 năm 2016

NHIỆM VỤ LUẬN VĂN THẠC SĨ

Họ tên học viên: NGUYỄN THỊ THU HIỀN

Giới tính: NỮ

Ngày, tháng, năm sinh: 06/11/1983

Nơi sinh: TP HCM

Chuyên ngành: Công nghệ thông tin

MSHV: 1441860035

I- Tên đề tài:

DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN DỰA TRÊN HỆ SUY LUẬN ANFIS.

II- Nhiệm vụ và nội dung:

Nghiên cứu hệ suy luận ANFIS để dự báo dữ liệu chuỗi thời gian.

III- Ngày giao nhiệm vụ: 23/01/2016

IV- Ngày hoàn thành nhiệm vụ: 30/06/2016

V- Cán bộ hướng dẫn: PGS. TS. QUẢN THÀNH THƠ

CÁN BỘ HƯỚNG DẪN

KHOA QUẢN LÝ CHUYÊN NGÀNH

PGS. TS. QUẢN THÀNH THƠ

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu, kết quả nêu trong Luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện Luận văn này đã được cảm ơn và các thông tin trích dẫn trong Luận văn đã được chỉ rõ nguồn gốc.

Học viên thực hiện Luận văn

Nguyễn Thị Thu Hiền

LỜI CẢM ƠN

Đầu tiên, tôi xin bày tỏ lòng biết ơn chân thành và sâu sắc nhất tới PGS.TS. Quán Thành Thơ. Thầy đã tận tình hướng dẫn, động viên và giúp đỡ tôi thực hiện đề tài nghiên cứu này.

Tôi xin gửi lời cảm ơn chân thành đến PGS.TSKH. Nguyễn Xuân Huy, PGS.TS. Võ Đình Bảy, PGS.TS. Đỗ Phúc, TS. Nguyễn An Khương, PGS.TS. Lê Hoàng Thái, TS. Lư Nhật Vinh đã tận tình hướng dẫn, trang bị cho tôi những kiến thức nền tảng vô cùng quý báu trong suốt thời gian học tập tại trường.

Xin gửi lời cảm ơn đến cán bộ phòng Quản lý khoa học và đào tạo sau đại học của trường Đại học Công nghệ Tp.HCM đã hỗ trợ học vụ cho tôi trong thời gian theo học tại trường.

Xin gửi lời cảm ơn đến con gái yêu dấu đã động viên, tiếp thêm sức mạnh tinh thần cho mẹ trong suốt quá trình học tập và nghiên cứu.

Xin gửi lời cảm ơn đến ba mẹ và gia đình đã hỗ trợ, ủng hộ cho con trong suốt thời gian thực hiện luận văn.

Sau cùng, tôi xin gửi lời cảm ơn đến bạn bè và đồng nghiệp đã quan tâm, động viên và tạo mọi điều kiện để tôi hoàn thành luận văn này.

Nguyễn Thị Thu Hiền

TÓM TẮT

Khai thác dữ liệu quá khứ để dự báo dữ liệu tương lai với độ chính xác cao là vấn đề thử thách. Hệ suy luận ANFIS (*Adaptive Neuro-fuzzy Inference System*) là một trong các kỹ thuật dùng để dự báo dữ liệu chuỗi thời gian. Vì vậy, luận văn đề xuất mô hình dự báo đa trị dựa trên hệ suy luận ANFIS. Từ đó, luận văn xây dựng thuật toán huấn luyện và thuật toán kiểm thử dựa trên mô hình đã đề xuất. Để minh chứng tính hiệu quả của mô hình đề xuất, luận văn xây dựng ứng dụng thực nghiệm dự báo giá cổ phiếu trên cơ sở dữ liệu Cophieu68. Để thực hiện bài toán dự báo giá cổ phiếu dựa trên mô hình mạng ANFIS, luận văn lần lượt tiếp cận các phương pháp suy luận mờ, mạng nơron nhân tạo ANN và cấu trúc mô hình mạng neuro-fuzzy truyền thống. Từ đó, luận văn đề xuất mạng neuro-fuzzy ANFIS và mô hình ứng dụng thực nghiệm cho bài toán dự báo giá cổ phiếu trên cơ sở mô hình đã đề xuất, luận văn trình bày các thuật toán huấn luyện và kiểm thử dựa trên ngôn ngữ Matlab. Trong phần ứng dụng thực nghiệm, luận văn sẽ thực hiện một ví dụ thực nghiệm cụ thể để minh họa rõ về quá trình hoạt động của mạng neuro-fuzzy ANFIS đã đề xuất. Ngoài ra, trong phần ứng dụng thực nghiệm này đánh giá tính hiệu quả thông qua thời gian thực thi cũng như đánh giá độ chính xác thông qua việc so sánh với các phương pháp khác để từ đó cho thấy tính khả thi của phương pháp.

ABSTRACT

Prediction of data trend in future is one of major data mining tasks, which poses much challenges in order to obtain results of high accuracy. In particular, for the domain of time-series data, the inference system ANFIS (*Adaptive Neuro-fuzzy Inference System*) promises a powerful technique. Thus, this thesis proposes an approach of ANFIS-based multi-valued model. Further, the works carried on include developments of training algorithm and testing algorithm. To illustrate the efficiency of the proposed model, the thesis demonstrates an application of stock price prediction using Cophieu68 database. To achieve this goal, in this thesis the fuzzy inference, ANN technique and the neuro-fuzzy system are implemented. Eventually, the neuro-fuzzy ANFIS system is developed and applied for the application of stock price prediction, whose training and testing algorithms are developed using Matlab. In the experiments, the thesis also presents some case-studies to clearly illustrate the operation of the proposed neuro-fuzzy ANFIS system. In addition, we also evaluate the system performance on the metrics of execution time and accuracy, as well as comparison on other methods.

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
TÓM TẮT	iii
ABSTRACT	iv
MỤC LỤC.....	v
DANH MỤC TỪ VIẾT TẮT.....	viii
DANH MỤC CÁC BẢNG.....	ix
DANH MỤC HÌNH VẼ, ĐỒ THỊ	x
CHƯƠNG 1 TỔNG QUAN VỀ DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN	1
1.1. Giới thiệu	1
1.2. Mục đích và ý nghĩa đề tài	1
1.3. Khảo sát các công trình liên quan	3
1.4. Định hướng nghiên cứu	5
1.5. Các đối tượng nghiên cứu	5
1.5.1. Dữ liệu chuỗi thời gian	5
1.5.2. Mạng nơron ANN	6
1.5.3. Hệ suy luận mờ Takagi-Sugeno	7
1.5.4. Mạng hệ suy luận mờ ANFIS	8
1.6. Các nội dung cần nghiên cứu	9
1.7. Tổng kết chương.....	11
CHƯƠNG 2 DỮ LIỆU CHUỖI THỜI GIAN	12
2.1. Giới thiệu.....	12
2.2. Dữ liệu chuỗi thời gian.....	12
2.2.1. Khái niệm dữ liệu chuỗi thời gian	12
2.2.2. Biểu diễn dữ liệu chuỗi thời gian.....	12
2.3. Các kỹ thuật khai thác dự báo dữ liệu chuỗi thời gian.....	15
2.3.1. Phân lớp	15
2.3.2. Chỉ mục.....	15

2.3.3. Gom cụm	17
2.3.4. Dự báo.....	18
2.3.5. Tổng hợp.....	18
2.3.6. Phát hiện bất thường	20
2.3.7. Phân đoạn.....	21
2.4. Phân tích và chọn lựa các kỹ thuật.....	22
2.5. Tổng kết chương	22
CHƯƠNG 3 CÁC LÝ THUYẾT CƠ SỞ XÂY DỰNG HỆ SUY LUẬN ANFIS ..	23
3.1. Giới thiệu.....	23
3.2. Hệ suy luận mờ.....	23
3.2.1. Tổng quan về suy luận mờ.....	23
3.2.2. Logic mờ.....	24
3.2.3. Tập mờ và biến ngôn ngữ	25
3.2.4. Biểu diễn tập mờ.....	27
3.2.5. Các phương pháp khử mờ.....	30
3.2.6. Lập luận xấp xỉ mờ	33
3.3. Mạng nơron nhân tạo.....	34
3.3.1. Giới thiệu về mạng nơron nhân tạo	34
3.3.2. Hoạt động của mạng nơron nhân tạo	36
3.3.3. Một số kỹ thuật huấn luyện mạng	43
3.3.4. Dự báo dữ liệu chuỗi thời gian	44
3.4. Tổng kết chương.....	44
CHƯƠNG 4 HỆ SUY LUẬN ANFIS	45
4.1. Giới thiệu	45
4.2. Xây dựng mô hình suy luận mờ ANFIS nhằm dự báo giá cổ phiếu	45
4.4.1. Mô tả dữ liệu đầu vào	45
4.4.2. Xây dựng mô hình ANFIS.....	46
4.4.3. Thuật toán huấn luyện hệ suy luận mờ ANFIS	48
4.4.4. Thuật toán kiểm thử hệ suy luận mờ ANFIS.....	51

4.3. Ví dụ thực nghiệm	52
4.4. Tổng kết chương.....	63
CHƯƠNG 5 XÂY DỰNG ỨNG DỤNG THỰC NGHIỆM	64
5.1. Giới thiệu	64
5.2. Mô tả dữ liệu thực nghiệm	64
5.3. Mô tả mô hình thực nghiệm	66
5.4. Quá trình thực nghiệm.....	68
5.4.1. Các hàm chức năng chính.....	68
5.4.2. Thực nghiệm pha huấn luyện	69
5.4.3. Thực nghiệm pha kiểm thử.....	71
5.5. Kết quả thực nghiệm và đánh giá kết quả	72
5.5.1. Mô tả kết quả thực nghiệm	72
5.5.2. So sánh kết quả	76
5.6. Kết luận	78
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	79
DANH MỤC CÔNG TRÌNH LIÊN QUAN CỦA TÁC GIẢ ĐÃ CÔNG BỐ	80
TÀI LIỆU THAM KHẢO.....	81

DANH MỤC TỪ VIẾT TẮT

Ký hiệu	Tiếng Anh	Tiếng Việt
ANFIS	Adaptive Neuro-fuzzy Inference System	
MAE	Mean Absolute Error	Trung bình độ lệch
MAPE	Mean Absolute Percentage Error	Tỉ lệ phần trăm độ lệch
MPE	Mean Percentage Error	Trung bình phần trăm độ lệch
MSE	Mean Squared Error	Trung bình bình phương độ lệch
RMSE	Root Of The Mean Squared Error	Căn bậc hai của trung bình bình phương độ lệch

DANH MỤC CÁC BẢNG

Bảng 4.1. Ví dụ dữ liệu mẫu huấn luyện giá cổ phiếu	52
Bảng 4.2. Giá trị tham số mờ hàm Gaussian.....	53
Bảng 4.3. Kết quả mờ hóa.....	53
Bảng 4.4. Các giá trị trọng số mờ	55
Bảng 4.5. Chuẩn hóa các giá trị trọng số mờ	55
Bảng 4.6. Ma trận hệ số suy luận mờ lần huấn luyện thứ 1	57
Bảng 4.7. Các giá trị dự báo $f_i^{(N)}$ lần huấn luyện thứ 1	57
Bảng 4.8. Chuẩn hóa các giá trị dự báo lần huấn luyện thứ 1	58
Bảng 4.9. Kết quả dự báo lần huấn luyện thứ 1	60
Bảng 4.10. Ma trận hệ số suy luận mờ lần huấn luyện thứ 8.....	61
Bảng 4.11. Các giá trị dự báo $f_i^{(N)}$ lần huấn luyện thứ 8.....	61
Bảng 4.12. Chuẩn hóa các giá trị dự báo lần huấn luyện thứ 8	62
Bảng 4.13. Kết quả dự báo lần huấn luyện thứ 8.....	62
Bảng 5.1. Đánh giá sai số huấn luyện bộ dữ liệu từ M_1 đến M_{50}	73
Bảng 5.2. Đánh giá độ lệch giữa giá trị dự báo và giá trị thực tế	75
Bảng 5.3. So sánh sai số giữa phương pháp đề xuất và các phương pháp khác	76
Bảng 5.4. So sánh tỉ lệ sai số giữa phương pháp đề xuất và phương pháp ANN	77
Bảng 5.5. So sánh sai số RMSE giữa phương pháp đề xuất và phương pháp mạng nơron mờ loại 2	77
Bảng 5.6. So sánh sai số giữa phương pháp đề xuất và các phương pháp khác	78

DANH MỤC HÌNH VẼ, ĐỒ THỊ

Hình 1.1. Minh họa về sự biến thiên của dữ liệu theo thời gian	6
Hình 1.2. Mô tả mạng nơron truyền thẳng 3 lớp.....	6
Hình 1.3. Minh họa quá trình suy luận mờ	7
Hình 1.4. Minh họa mô hình mạng ANFIS.....	8
Hình 2.1. Giảm số chiều của chuỗi thời gian bằng cách lấy mẫu	13
Hình 2.2. Giảm số chiều của chuỗi thời gian bằng PAAA	13
Hình 2.3. Giảm chuỗi thời gian bằng cách giữ lại các điểm quan trọng.....	14
Hình 2.4. Giảm số chiều của dữ liệu chuỗi thời gian thành 2 chiều	16
Hình 2.5. Cấu trúc chỉ mục phân cấp sử dụng cây R-tree.....	16
Hình 2.6. Gom cụm phân cấp của chuỗi thời gian.....	17
Hình 2.7. Một ví dụ về trừu tượng hóa lịch và gom cụm.....	19
Hình 2.8. Ví dụ về sử dụng điện năng hàng năm.....	20
Hình 2.9. Minh họa các bất thường trên dữ liệu ECG	20
Hình 2.10. Một phát hiện bất thường từ dữ liệu kiểm thử MIT-BIH.....	21
Hình 2.11. Một phân đoạn chuỗi thời gian	21
Hình 3.1. Mô hình lập luận xấp xỉ mờ trên giá trị ngôn ngữ	27
Hình 3.2. Biểu diễn tập mờ hàm mũ ứng với giá trị chân lý TRUE	28
Hình 3.3. Biểu diễn tập mờ hàm mũ ứng với giá trị chân lý FALSE	28
Hình 3.4. Biểu diễn tập mờ tam giác ứng với giá trị chân lý TRUE	29
Hình 3.5. Biểu diễn tập mờ tam giác ứng với giá trị chân lý FALSE.....	29
Hình 3.6. Biểu diễn tập mờ tuyến tính với giá trị chân lý TRUE	30
Hình 3.7. Biểu diễn tập mờ tuyến tính với giá trị chân lý FALSE	30
Hình 3.8. Minh họa mô hình mạng nơron ANN 4 lớp.....	35
Hình 3.9. Mô tả hàm tính tổng trọng.....	37
Hình 3.10. Mô tả hàm truyền sigmoid	37
Hình 3.11. Hàm truyền tan- hyperbol	38
Hình 4.1. Mô hình ANFIS dự báo giá cổ phiếu	48
Hình 5.1. Biểu đồ dữ liệu huấn luyện giá cổ phiếu (ĐVT: 1,000VNĐ).....	64

Hình 5.2. Biểu đồ dữ liệu mờ hóa của tập dữ liệu huấn luyện.....	65
Hình 5.3. Hàm mờ hóa tập dữ liệu huấn luyện	65
Hình 5.4. Mô tả dữ liệu kiểm thử giá cổ phiếu (ĐVT: 1,000VNĐ)	66
Hình 5.5. Mô hình thực nghiệm dự báo giá cổ phiếu	67
Hình 5.6. Sai số huấn luyện giá cổ phiếu với trạng thái <Open>	72
Hình 5.7. Sai số huấn luyện giá cổ phiếu với trạng thái <High>.....	72
Hình 5.8. Sai số huấn luyện giá cổ phiếu với trạng thái <Low>	73
Hình 5.9. Sai số huấn luyện giá cổ phiếu với trạng thái <Close>.....	73
Hình 5.10. Thời gian huấn luyện của toàn bộ tập dữ liệu	74
Hình 5.11. Mô tả giá trị dự báo và giá trị thực.....	74
Hình 5.12. Độ lệch trung bình giữa giá trị dự báo và giá trị thực tế.....	75
Hình 5.13. Đồ thị mô tả thời gian dự báo	75

CHƯƠNG 1

TỔNG QUAN VỀ DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN

1.1. Giới thiệu

Trong chương này tiếp cận nghiên cứu về bài toán dự báo dữ liệu chuỗi thời gian. Để thực hiện vấn đề này, nội dung của chương trình bày các công trình nghiên cứu liên quan đến dự báo dữ liệu chuỗi thời gian, đặc biệt là các công trình nghiên cứu về dự báo giá cổ phiếu để từ đó luận văn đề xuất mô hình dự báo giá cổ phiếu dựa trên dữ liệu Cophieu68.

Nội dung của chương lần lượt tiếp cận các đối tượng nghiên cứu liên quan bao gồm: (1) Dữ liệu chuỗi thời gian; (2) Mạng nơron nhân tạo ANN; (3) Hệ suy luận mờ; (4) Mạng suy luận neuro-fuzzy ANFIS.

1.2. Mục đích và ý nghĩa đề tài

Theo thống kê của Tufte [14], hầu hết nguồn cung cấp dữ liệu của thế giới là ở dạng chuỗi thời gian, có khoảng 75% dữ liệu hình ảnh trên các tờ báo ở dạng chuỗi thời gian và kích thước của dữ liệu chuỗi thời gian tăng theo cấp số nhân. Bài toán dự báo dữ liệu chuỗi thời gian được ứng dụng trong nhiều lĩnh vực như dự báo giá cổ phiếu, dự báo thời tiết, dự báo sản lượng sản xuất,... [3, 27]. Kết quả thu được từ mô hình dự báo có thể trở thành nguồn dữ liệu tư vấn và có thể áp dụng cho các loại dữ liệu khác nhau, từ đó đóng góp một công cụ tư vấn hữu ích cho cộng đồng.

Theo như F.Zang và cộng sự [33], bài toán dự báo giá cổ phiếu có liên hệ mật thiết với nền kinh tế quốc nội, có sự biến đổi phức tạp và phi tuyến trong một khoảng thời gian ngắn. Bên cạnh đó, giá của cổ phiếu trong thị trường cổ phiếu thường liên quan đến các dữ liệu quá khứ. Trong công trình này, F.Zang đã khảo sát các phương pháp truyền thống dựa trên các kỹ thuật phân tích tuyến tính sẽ rất khó tiếp cận bài toán dự báo giá thị trường cổ phiếu. Vì vậy, cần phải có một kỹ thuật thông minh để giải quyết bài toán này. Từ đó, góp một phần ý nghĩa trong nền kinh tế hiện đại, đặc biệt trong luận văn mong muốn đóng góp một công cụ hữu ích cho thị trường cổ phiếu tại Việt Nam.

Theo như Thangjam Ravichandra và cộng sự [19], thị trường tài chính là một trong những thị trường phức tạp nhất và rất khó xác định quá trình biến đổi giá cổ phiếu bởi vì có rất nhiều yếu tố và nguyên nhân tác động. Vì vậy, việc dự báo giá cổ phiếu chính xác là một công việc khó khăn. Trong công trình này Thangjam Ravichandra đã mô tả nhiều phương pháp khai phá dữ liệu để dự báo giá cổ phiếu trong thị trường tài chính. Tuy nhiên, mỗi một phương pháp tiếp cận đều có những ưu khuyết điểm khác nhau. Hơn nữa, vấn đề quan trọng nhất đó là sai số của việc dự báo cổ phiếu, tức là sai số càng thấp thì phương pháp càng có giá trị. Trong công trình này cũng đã tiếp cận mạng ANN và minh chứng bằng thực nghiệm cho thấy phương pháp dự báo này hiệu quả. Vì vậy, trong luận văn này mong muốn đóng góp một công cụ dự báo giá cổ phiếu hiệu quả, tức là sai số thấp nhất có thể để từ đó góp một công cụ tư vấn tài chính cho xã hội.

Theo như T.Nguyen và cộng sự [18] việc dự báo giá cổ phiếu giúp cho các nhà đầu tư có một khung nhìn tổng thể và định hướng sự biến đổi giá cổ phiếu. Dựa trên kết quả dự báo, các nhà đầu tư có thể đưa ra quyết định để thực hiện việc mua hoặc bán tại một thời điểm thích hợp. Do đó, bài toán dự báo cổ phiếu giúp cho các nhà đầu tư giảm thiểu rủi ro tài chính. Vì vậy, luận văn cũng mong muốn các nhà đầu tư tại Việt Nam có thêm một công cụ hỗ trợ cho quá trình ra quyết định của chính mình.

Theo như Mu-Yen Chen và cộng sự [5], dữ liệu giá cổ phiếu là một dạng quan trọng trong khái niệm dữ liệu chuỗi thời gian. Theo như công trình này đã khảo sát có rất nhiều phương pháp dự báo dựa trên thống kê, các mô hình chuỗi thời gian, kỹ thuật logic mờ, kỹ thuật mạng nơron, các kỹ thuật dự báo qua khai phá dữ liệu,... Các kỹ thuật đã khảo sát cũng đều có những ưu khuyết khác nhau.

Vì vậy, trong luận văn tiếp cận một kỹ thuật mang tính chất kết hợp giữa suy luận mờ và mạng nơron nhân tạo để xây dựng một hệ ANFIS nhằm dự báo nhiều loại giá cổ phiếu khác nhau.

Luận văn mong muốn xây dựng một mô hình cải tiến mạng ANFIS truyền thống nhằm tạo ra một phương pháp hiệu quả nhằm dự báo đồng thời nhiều kết quả cho các nhà kinh tế.

1.3. Khảo sát các công trình liên quan

Có nhiều mô hình dự báo đã được công bố trong thời gian gần đây như sử dụng cấu trúc cây TAN (*Tree Augmented Naïve*) Bayes [13], mạng nơron ANN [5, 25, 35], mô hình SVM [8, 46],... Trong các mô hình dự báo, hệ ANFIS phù hợp cho các bài toán có dữ liệu đầu vào phức tạp và dự báo được đồng thời nhiều kết quả khác nhau, ví dụ như dự báo giá cổ phiếu dựa trên Hamacher T-Norm nhiều đầu vào và ANFIS [27], dự báo giá đóng (close) [23], mô hình ANFIS áp dụng trên di động [1].

Có nhiều bài toán dự báo dữ liệu chuỗi thời gian dựa trên mô hình mạng nơron như: dự báo nguồn tài nguyên nước [3], dự báo lượng điện cần tải cho thành phố [20], dự báo sự biến đổi của năng lượng mặt trời [22], dự báo mức độ ô nhiễm môi trường [17],...

Ngoài phương pháp sử dụng mô hình mạng nơron, một số bài toán áp dụng hệ thống suy luận mờ FIS (*Fuzzy Inference System*) để đưa ra các dự báo như: dự báo chỉ số chứng khoán [24], dự báo tỷ giá thông qua kỹ thuật tính toán DNA [9], dự báo sự biến động của giá cả thị trường cổ phiếu [11], hệ hỗ trợ ra quyết định chiến lược kinh doanh [6],...

Mô hình dự báo neuro-fuzzy ANFIS dựa trên sự kết hợp giữa mô hình mạng nơron và hệ suy luận mờ FIS cũng được đề cập trong nhiều công trình nghiên cứu như phân loại phương tiện và lưu lượng giao thông trong các ứng dụng vận chuyển [15], dự báo sự hỏng hóc của máy móc [25], xác định xác suất xảy ra dịch sốt xuất huyết ở Bangladesh [2], kỹ thuật mạng nơron kết hợp với suy luận mờ để dự báo giá cổ phiếu, nhiệt độ, bán hàng, sản xuất cây trồng [21], xây dựng mô hình F-CONFIS (*Fully Connected Neuro-Fuzzy Inference System*) để dự báo nhiều dạng dữ liệu đầu ra [4], xây dựng kiến trúc mạng neuro-fuzzy và thuật toán suy luận mờ để dự báo dữ liệu trong tương lai [10], kết hợp giữa logic mờ, mạng nơron, mô hình phân lớp dựa trên mạng neuro-fuzzy [12],...

Theo tài liệu [27], vào năm 2015 F. Zhang và cộng sự tiếp cận hệ ANFIS để dự báo giá cổ phiếu. Trong công trình này sử dụng các đặc tính của chuẩn T-Norm và hệ ANFIS để dự báo giá cổ phiếu. Trong thực nghiệm của công trình này đã đánh giá

các dạng sai số dự báo gồm trung bình bình phương độ lệch MSE (*Mean Squared Error*), trung bình độ lệch MAE (*Mean Absolute Error*) và tỉ lệ phần trăm độ lệch MAPE (*Mean Absolute Percentage Error*). Kết quả thực nghiệm của công trình này cho thấy tính hiệu quả của mạng ANFIS để dự báo giá cổ phiếu.

Theo tài liệu [19], vào năm 2016 Thangjam Ravichandra và cộng sự sử dụng mạng nơron nhân tạo ANN để dự báo giá cổ phiếu. Trong phương pháp của công trình này đã đề xuất hai phương pháp gồm mạng nơron 3 lớp và phương pháp mạng nơron kết hợp với cây quyết định. Theo thực nghiệm cho thấy kết quả đánh giá tỉ lệ phần trăm hàm lỗi có giá trị thấp. Điều này cho thấy mạng ANN là một hướng tiếp cận phù hợp để dự báo giá cổ phiếu.

Theo tài liệu [18], vào năm 2013 T. Nguyen và cộng sự sử dụng mạng nơron và tập mờ loại 2 để dự báo giá cổ phiếu. Trong thực nghiệm đã đánh giá sai số dự báo căn bậc hai của trung bình bình phương độ lệch RMSE (*Root Of The Mean Squared Error*) nhằm đưa ra tính hiệu quả của phương pháp đã đề xuất trong công trình này.

Theo tài liệu [5], vào năm 2015 Mu-Yen Chen và cộng sự thực hiện công trình dự báo giá cổ phiếu dựa trên mô hình hệ lai bằng cách kết hợp các dạng tập mờ và độ biến thiên entropy. Trong công trình này, các tác giả đã đánh giá các dạng sai số gồm trung bình bình phương độ lệch MSE, căn bậc hai của trung bình bình phương độ lệch RMSE, trung bình độ lệch MAE và trung bình phần trăm độ lệch MPE (*Mean Percentage Error*). Theo như kết quả sai số thực nghiệm cho thấy phương pháp sử dụng nội suy mờ dựa trên hệ lai của bài báo này thể hiện tính hiệu quả cho bài toán dự báo cổ phiếu.

Từ các công trình trên cho thấy mô hình ANFIS phù hợp với bài toán dự báo dữ liệu chuỗi thời gian.

Trên cơ sở các công trình liên quan đã khảo sát, luận văn đề xuất mô hình dự báo giá cổ phiếu theo phương thức dự báo đồng thời nhiều kết quả dựa trên hệ ANFIS và hệ suy luận mờ Takagi-Sugeno.

Từ đó, luận văn trình bày các thuật toán huấn luyện và thuật toán kiểm thử tương ứng. Để minh chứng tính hiệu quả của mô hình đề xuất, luận văn xây dựng ứng dụng thực nghiệm trên dữ liệu thực tế từ Cophieu68 [34].

1.4. Định hướng nghiên cứu

Luận văn tiếp cận hệ suy luận mờ ANFIS nhằm dự báo giá cổ phiếu dựa trên dữ liệu thực tế từ Cophieu68, vì vậy nội dung nghiên cứu của luận văn như sau:

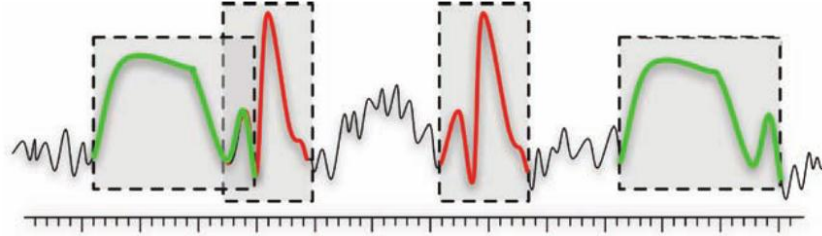
- Nghiên cứu về dữ liệu chuỗi thời gian nhằm xây dựng ứng dụng dự báo giá cổ phiếu dựa trên mô hình ANFIS.
- Nghiên cứu cơ sở lý thuyết về tập mờ và các phương pháp lập luận suy diễn mờ để từ đó kết hợp mô hình mạng neuro-fuzzy nhằm xây dựng hệ dự báo giá cổ phiếu.
- Nghiên cứu về mạng nơron nhân tạo ANN để làm cơ sở tạo ra mạng ANFIS nhằm giải quyết bài toán dự báo giá cổ phiếu.
- Nghiên cứu cấu trúc mạng ANFIS để từ đó đề xuất mô hình dự báo phù hợp với giá cổ phiếu dựa trên dữ liệu Cophieu68.
- Trên cơ sở lý thuyết, luận văn đưa ra thuật toán huấn luyện mạng ANFIS đã đề xuất cũng như thuật toán kiểm thử.
- Luận văn xây dựng ứng dụng thực nghiệm dự báo giá cổ phiếu nhằm minh chứng tính đúng đắn của cơ sở lý thuyết dựa trên công cụ lập trình Matlab. Từ đó, luận văn trình bày các kết quả so sánh với các phương pháp khác cùng thể loại để đánh giá tính hiệu quả của mô hình đề xuất.

1.5. Các đối tượng nghiên cứu

1.5.1. Dữ liệu chuỗi thời gian

Theo như các công trình liên quan đã khảo sát, dữ liệu giá cổ phiếu là một dạng dữ liệu chuỗi thời gian. Do đó, đối tượng nghiên cứu cần phải khảo sát đó là dữ liệu chuỗi thời gian. Vì vậy, luận văn tiến hành ánh xạ dữ liệu giá cổ phiếu để tạo thành dữ liệu dưới dạng chuỗi thời gian nhằm thực hiện chuẩn hóa dữ liệu đầu vào cho bài toán dự báo giá cổ phiếu trong tương lai. Trong cách tiếp cận của luận văn xem như dữ liệu chuỗi thời gian là đối tượng nghiên cứu ban đầu để chọn lựa các mô

hình dự báo phù hợp, để từ đó làm tiền đề xây dựng một công cụ dự báo giá cổ phiếu hiệu quả.

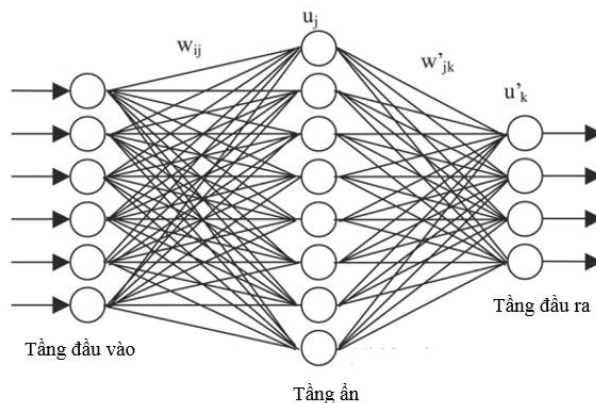


Hình 1.1. Minh họa về sự biến thiên của dữ liệu theo thời gian [7]

Trong **Hình 1.1** là một mô tả về sự biến thiên của các giá trị dữ liệu theo trục thời gian. Sự biến thiên này là phi tuyến và không có một quy luật cụ thể, do đó cần xây dựng một công cụ tích hợp với tri thức để tiên lượng của các giá trị này trong tương lai.

1.5.2. Mạng nơron ANN

Luận văn tiếp cận một công cụ dự báo giá cổ phiếu trên cơ sở mạng ANFIS. Do đó, mạng nơron ANN là một đối tượng nghiên cứu không thể thiếu bởi vì mạng ANFIS là sự kết hợp giữa mạng nơron nhân tạo và hệ suy luận mờ. Trong luận văn sẽ khảo sát mạng nơron nhân tạo ANN để từ đó làm cơ sở đề xuất mô hình dự báo cải tiến. Các phương pháp huấn luyện và phương pháp kiểm thử trên mô hình đề xuất của luận văn cũng có nguồn gốc từ mạng nơron nhân tạo ANN truyền thống. Vì vậy, đối tượng nghiên cứu thứ hai của luận văn đó là mạng nơron nhân tạo truyền thống, các tính chất của mạng nơron, các phương pháp huấn luyện, kiểm thử của mạng nơron cũng sẽ lần lượt được khảo sát.

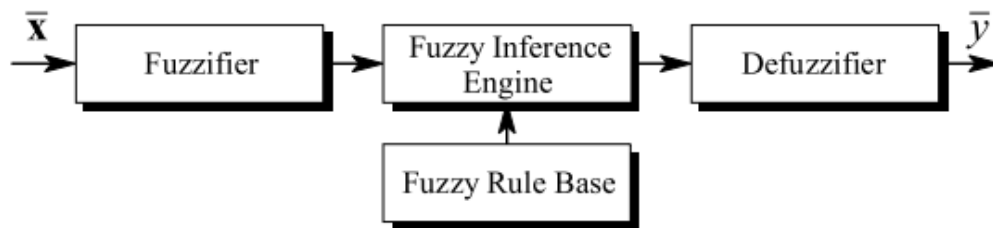


Hình 1.2. Mô tả mạng nơron truyền thẳng 3 lớp

1.5.3. Hệ suy luận mờ Takagi-Sugeno

Trong hệ suy luận ANFIS cần nội suy các giá trị đầu vào để tạo thành giá trị đầu ra. Các giá trị đầu vào thuộc về tập các giá trị rõ sau đó thực hiện mờ hóa nhằm áp dụng các phương pháp nội suy mờ để có được kết quả là giá trị mờ, từ đó thực hiện phương pháp khử mờ để tìm ra giá trị rõ đầu ra tương ứng.

Nhằm giảm thời gian khử mờ, trong luận văn áp dụng phương pháp suy luận mờ Takagi-Sugeno để tìm trực tiếp giá trị rõ đầu ra mà bỏ qua công đoạn khử mờ. Vì vậy, phương pháp suy luận mờ Takagi-Sugeno là một trong những đối tượng nghiên cứu chính của luận văn để từ đó áp dụng phương pháp này để nội suy kết quả đầu ra.



Hình 1.3. Minh họa quá trình suy luận mờ [16]

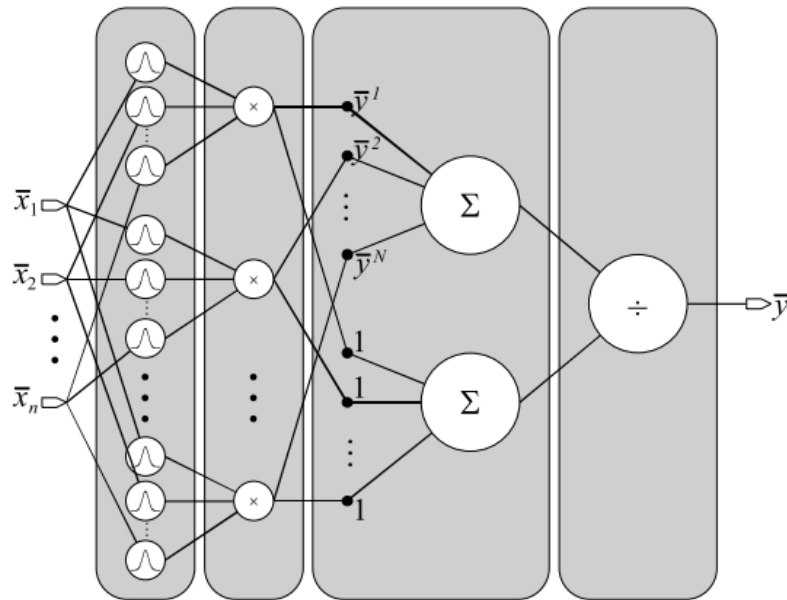
Trong **Hình 1.3** mô tả quá trình suy luận mờ cơ bản bao gồm giá trị đầu vào, sau đó thực hiện mờ hóa giá trị đầu vào. Bước quan trọng tiếp theo đó là thực hiện cơ chế suy luận mờ dựa trên cơ sở luật mờ để tạo ra giá trị mờ đầu ra.

Bước sau cùng của hệ thống suy luận mờ này là thực hiện quá trình khử mờ để tìm giá trị rõ đầu ra tương ứng. Trong luận văn áp dụng phương pháp suy luận mờ Takagi-Sugeno, nên tìm ra trực tiếp các giá trị rõ đầu ra mà không cần giai đoạn khử mờ. Phương pháp suy luận mờ này về cơ bản tuân thủ các quy tắc của mô hình suy luận mờ, nhưng bỏ qua công đoạn khử mờ của kết quả đầu ra bởi vì tìm được trực tiếp kết quả của giá trị rõ.

Tuy nhiên, phương pháp Takagi-Sugeno cần phải tìm ra bảng hệ số suy luận mờ để làm tiền đề cho phương pháp suy luận nhằm có được kết quả đầu ra. Việc tìm hệ số suy luận này được thực hiện dựa trên quá trình huấn luyện mạng neuro-fuzzy ANFIS trên cơ sở phương pháp huấn luyện mạng nơron nhân tạo ANN.

1.5.4. Mạng hệ suy luận mờ ANFIS

Luận văn tiếp cận bài toán dự báo giá cổ phiếu từ dữ liệu quá khứ bằng phương pháp suy luận mờ dựa trên mạng neuro-fuzzy ANFIS. Trong cách tiếp cận này, luận văn thực hiện cải tiến mô hình mạng neuro-fuzzy ANFIS nhằm phù hợp với dữ liệu giá chứng khoán Cophieu68 dưới dạng dữ liệu chuỗi thời gian. Để thực hiện được vấn đề này, đối tượng nghiên cứu quan trọng nhất đó là mô hình mạng ANFIS truyền thống để từ đó xây dựng mô hình cải tiến và thiết kế mô hình ứng dụng nhằm tạo ra một công cụ dự báo giá cổ phiếu cho cộng đồng.



Hình 1.4. Minh họa mô hình mạng ANFIS [16]

Trong **Hình 1.4** mô tả một mạng neuro-fuzzy ANFIS cơ sở bao gồm một lớp đầu vào là một vector các giá trị rõ. Sau đó, dựa trên mô hình này thực hiện quá trình mờ hóa và tạo ra các trọng số mờ tương ứng để từ đó tổng hợp dựa trên các hệ số suy luận mờ nhằm tạo ra giá trị rõ đầu ra.

Quá trình mờ hóa được thực hiện có thể bằng nhiều phương pháp mờ hóa khác nhau như dùng tập mờ tam giác, tập mờ hình thang, tập mờ dạng hình chuông,... Việc suy luận mờ để tạo kết quả đầu ra được ứng dụng phương pháp suy luận mờ Takagi-Sugeno. Nếu giá trị suy luận mờ này so với giá trị thực tế có hàm lỗi chưa thỏa mãn điều kiện hội tụ thì quá trình này thực hiện quay lui nhằm huấn luyện mạng neuro-

fuzzy để tinh chỉnh tập các giá trị trọng số mờ nhằm tìm ra giá trị đầu ra gần với giá trị thực tế nhiều nhất có thể. Sau khi quá trình huấn luyện hội tụ, mạng neuro-fuzzy đã có được tập các hệ số suy luận mờ tương ứng. Trên cơ sở các hệ số suy luận mờ này, mạng neuro-fuzzy áp dụng cho quá trình kiểm thử để có được kết quả đầu ra tương ứng.

1.6. Các nội dung cần nghiên cứu

Trên cơ sở các đối tượng nghiên cứu và các công trình liên quan đã được phân tích như trên, nội dung của luận văn được thiết kế gồm 5 chương như sau:

Chương 1. TỔNG QUAN VỀ DỰ BÁO DỮ LIỆU CHUỖI THỜI GIAN:

Trong chương này sẽ khảo sát tổng quan về bài toán dự báo dữ liệu chuỗi thời gian đặc biệt là bài toán dự báo giá cổ phiếu. Để thấy được mục đích, ý nghĩa và cơ sở khoa học của phương pháp giải quyết bài toán này, nội dung của chương lần lượt khảo sát các công trình liên quan cũng như các ý nghĩa xã hội của phương pháp giải quyết bài toán. Từ đó, trong chương này lần lượt tiếp cận các đối tượng nghiên cứu để làm nền tảng cho quá trình xây dựng luận văn. Trên cơ sở này, luận văn tiếp cận phương pháp nghiên cứu dựa trên cơ sở lý thuyết đã có qua các đối tượng nghiên cứu đã khảo sát. Bên cạnh đó, luận văn thực hiện phương pháp nghiên cứu bằng phương pháp thực nghiệm để minh chứng tính đúng đắn của cơ sở lý thuyết đã đề nghị.

Chương 2. DỮ LIỆU CHUỖI THỜI GIAN: Trong chương này sẽ khảo sát tổng quan về dữ liệu chuỗi thời gian nhằm áp dụng trong trường hợp cụ thể đó là dữ liệu giá cổ phiếu trên cơ sở dữ liệu Cophieu68. Vì vậy, nội dung chương này sẽ khảo sát các tính chất tổng quan và các đặc trưng của dữ liệu chuỗi thời gian để từ đó áp dụng vào các trường hợp cụ thể. Ngoài ra, trong chương này cũng khảo sát các kỹ thuật phân tích thao tác xử lý dạng dữ liệu chuỗi thời gian để từ đó có thể áp dụng làm dữ liệu đầu vào cho mạng neuro-fuzzy ANFIS nhằm xây dựng hệ dự báo giá cổ phiếu trong tương lai.

Chương 3. CÁC LÝ THUYẾT CƠ SỞ XÂY DỰNG HỆ SUY LUẬN ANFIS:

Trong chương này tiếp cận hai đối tượng nghiên cứu quan trọng đó là hệ suy luận mờ và mạng nơron nhân tạo. Đây là hai đối tượng cơ bản để cấu thành mạng neuro-fuzzy

ANFIS để từ đó làm cơ sở xây dựng hệ dự báo giá cổ phiếu. Vì vậy, nội dung của chương này sẽ lần lượt khảo sát các đặc tính và các đối tượng của hệ suy luận mờ. Các giá trị mờ, tập mờ, phương pháp mờ hóa, phương pháp khử mờ,... sẽ được lần lượt khảo sát nhằm áp dụng trong hệ suy luận mờ. Bên cạnh đó, phương pháp suy luận mờ cũng sẽ được khảo sát và phân tích nhằm lựa chọn phương pháp suy luận mờ phù hợp. Song song với hệ suy diễn mờ đã được khảo sát, trong chương này tiếp cận mạng nơron nhân tạo để làm cơ sở cho quá trình huấn luyện và kiểm thử của mạng neuro-fuzzy ANFIS đã được đề xuất. Vì vậy, các tính chất đặc trưng, các thuộc tính và các đối tượng liên quan đến mạng nơron ANN cũng sẽ được khảo sát để từ đó ứng dụng xây dựng mô hình suy luận neuro-fuzzy ANFIS, đồng thời xây dựng ứng dụng dự báo giá cổ phiếu.

Chương 4. HỆ SUY LUẬN ANFIS: Dựa trên cơ sở lý thuyết và các đối tượng nghiên cứu đã được tiếp cận, trong chương này đề xuất mô hình mạng neuro-fuzzy ANFIS cho bài toán dự báo cổ phiếu. Mô hình đề xuất trong chương này dựa trên cơ sở các mô hình và các đối tượng đã có, tuy nhiên mô hình đề xuất sẽ phù hợp với dữ liệu đầu vào là một ma trận các giá trị rõ và đầu ra là một vector mà mỗi thành phần là giá tương ứng với từng mã cổ phiếu. Trong chương này cũng sẽ mô tả rõ từng thành phần của mô hình đã đề xuất cũng như phân tích sự phù hợp ứng với bài toán dự báo cổ phiếu. Dựa trên mô hình đã đề xuất, trong chương này đề xuất hai thuật toán quan trọng bao gồm thuật toán huấn luyện và thuật toán kiểm thử nhằm tạo ra một mạng ANFIS để thực hiện pha dự báo cho người dùng. Nhằm làm rõ quá trình thực thi của mạng neuro-fuzzy ANFIS, trong chương này thực thi một ứng dụng cụ thể với các dữ liệu thực tế để cho thấy các giá trị của mỗi tầng trong mạng neuro-fuzzy ANFIS đã đề xuất. Nội dung của chương 4 này là tiền đề cơ sở để xây dựng ứng dụng thực nghiệm được mô tả trong chương 5.

Chương 5. XÂY DỰNG ỨNG DỤNG THỰC NGHIỆM: Nhằm minh chứng tính đúng đắn của cơ sở lý thuyết đã đề nghị, trong chương này xây dựng ứng dụng thực nghiệm và đánh giá các kết quả dựa trên dữ liệu thực tế của Cophieu68. Để thực hiện vấn đề này, nội dung của chương trình bày mô hình ứng dụng và các bước thực

thi ứng dụng nhằm xây dựng hệ dự báo giá cổ phiếu. Dựa trên mô hình ứng dụng đã xây dựng, các thành phần cơ sở của ứng dụng được trình bày nhằm minh chứng sự phù hợp của mô hình đã đề xuất, đồng thời các hàm chức năng chính dựa trên cơ sở thuật toán cũng sẽ được trình bày và làm rõ sau khi thực hiện ứng dụng, trong chương này cũng trình bày các kết quả thực nghiệm của quá trình huấn luyện và kiểm thử giá cổ phiếu. Hơn nữa, để minh chứng tính hiệu quả của phương pháp đề xuất, trong chương này thực hiện đánh giá và so sánh kết quả thực nghiệm với các công trình đồng dạng liên quan.

1.7. Tổng kết chương

Trong chương này đã khảo sát tổng quan về hệ dự báo giá cổ phiếu dựa trên mạng neuro-fuzzy ANFIS. Các đối tượng nghiên cứu và các công trình nghiên cứu liên quan cũng được tiếp cận từ đó cho thấy tính khả thi của phương pháp tiếp cận trong việc giải quyết bài toán này. Ngoài ra, trong chương này cũng đã khảo sát tổng quan các đối tượng cấu thành mạng neuro-fuzzy ANFIS nhằm làm tiền đề cơ sở xây dựng một mô hình cải tiến phù hợp với dữ liệu Cophieu68. Nhằm định hướng rõ về quá trình thực hiện luận văn, trong chương này lần lượt tiếp cận nội dung tổng quát về từng chương của luận văn để từ đó dễ dàng tiếp cận các đối tượng chi tiết trong hệ neuro-fuzzy ANFIS. Dựa trên cơ sở lý thuyết đã tiếp cận trong chương này, các nội dung về xây dựng hệ dự báo cổ phiếu sẽ được lần lượt trình bày từ Chương 2 đến Chương 5.

CHƯƠNG 2

DỮ LIỆU CHUỖI THỜI GIAN

2.1. Giới thiệu

Trong chương này trình bày các cơ sở lý thuyết về dữ liệu chuỗi thời gian như khái niệm dữ liệu chuỗi thời gian, cách biểu diễn dữ liệu chuỗi thời gian và các kỹ thuật khai thác dữ liệu chuỗi thời gian. Từ các kỹ thuật khai thác đó, luận văn tiến hành phân tích và lựa chọn kỹ thuật khai thác dữ liệu phù hợp với yêu cầu luận văn.

2.2. Dữ liệu chuỗi thời gian

Phần lớn nguồn cung cấp dữ liệu của thế giới là ở dạng chuỗi thời gian. Kích thước của dữ liệu chuỗi thời gian tăng theo cấp số nhân. Chẳng hạn như trong lĩnh vực y tế, các dữ liệu về gen, điện tâm đồ, điện não đồ, phân tích đáng đi, bảng xếp hạng tăng trưởng chiều cao liên tục được tạo ra. Bên cạnh đó, dữ liệu của các ngành công nghiệp, giải trí, tài chính, khí tượng và hầu hết các lĩnh vực của đời sống con người cũng liên tục được hình thành [14]. Đây chính là điều đã dẫn đến việc bùng nổ về khai thác dữ liệu chuỗi thời gian nhằm đưa ra các dự báo hữu ích phục vụ cho cộng đồng như dự báo thời tiết, dự báo giá cổ phiếu, dự báo sản lượng kinh doanh,....

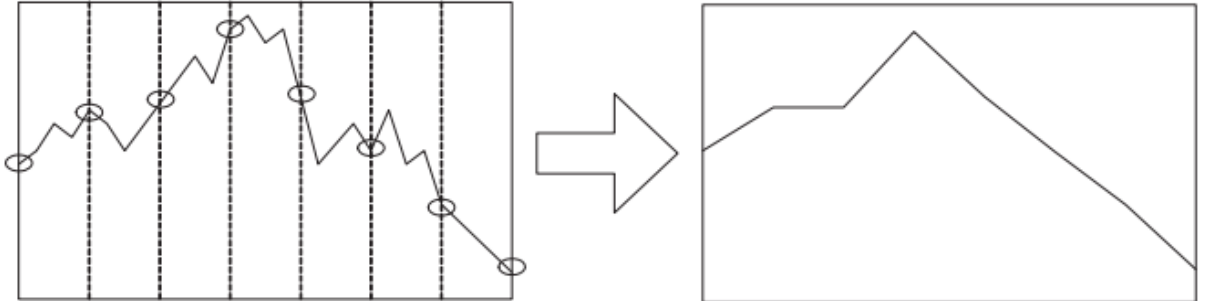
2.2.1. Khái niệm dữ liệu chuỗi thời gian

Một chuỗi thời gian là một tập hợp các quan sát được ghi nhận tại thời điểm cụ thể t . Chuỗi thời gian được phân ra thành hai loại: (1) Chuỗi thời gian rời rạc là tập các thời điểm quan sát được thực hiện tại các khoảng thời gian cố định; (2) Chuỗi thời gian liên tục là tập các thời điểm quan sát được ghi nhận liên tục qua một số khoảng thời gian. Một chuỗi thời gian T là một dãy có thứ tự của n giá trị thực $T = (t_1, t_2, \dots, t_n)$ với $t_i \in R$ [26].

2.2.2. Biểu diễn dữ liệu chuỗi thời gian

Một trong những lý do chủ yếu của việc biểu diễn chuỗi thời gian là nhằm giảm kích thước của dữ liệu gốc. Phương pháp đơn giản nhất là lấy mẫu. Trong phương pháp này, một tỷ lệ m/n được sử dụng với m là chiều dài của chuỗi dữ liệu P và n là số chiều sau khi giảm kích thước (**Hình 2.1**).

Tuy nhiên, phương pháp này có nhược điểm là bóp méo hình dạng của chuỗi thời gian khi tỉ lệ lấy mẫu quá thấp.

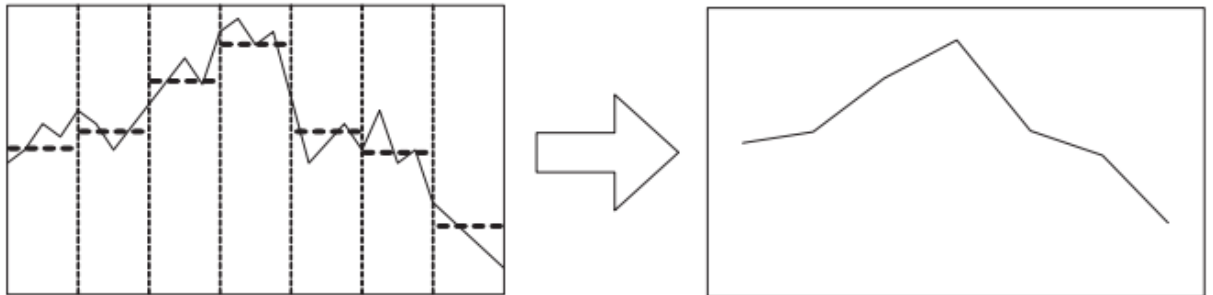


Hình 2.1. Giảm số chiều của chuỗi thời gian bằng cách lấy mẫu [8]

Phương pháp PAAA (*Piecewise Aggregate Approximation*) sử dụng các giá trị trung bình của mỗi đoạn để biểu diễn tập điểm dữ liệu tương ứng. Với chuỗi thời gian $P = (p_1, \dots, p_n)$ và n là số chiều đã giảm kích thước.

Chuỗi thời gian nén $\hat{P} = (\hat{p}_1, \dots, \hat{p}_n)$ có thể chứa bởi $\hat{p}_k = \frac{1}{e_k - s_k + 1} \sum_{i=s_k}^{e_k} p_i$ với

s_k là ký hiệu cho các điểm dữ liệu bắt đầu, e_k là ký hiệu cho các điểm dữ liệu kết thúc của phân đoạn thứ k trong chuỗi thời gian P (**Hình 2.2**).



Hình 2.2. Giảm số chiều của chuỗi thời gian bằng PAAA [8]

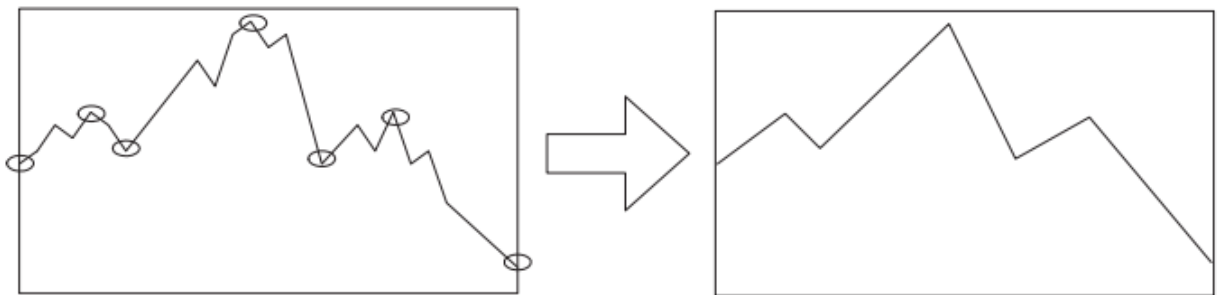
Một đề xuất mở rộng của phương pháp PAAA là APCA (*Adaptive Piecewise Constant Approximation*). Với phương pháp này thì chiều dài của mỗi phân đoạn là không cố định nhưng phù hợp với hình dạng của chuỗi.

Nhằm giảm số chiều của dữ liệu chuỗi, phương pháp PLR (*Piecewise Linear Representation*) đề xuất xấp xỉ chuỗi thời gian tương đương với một đường thẳng.

Đường thẳng xấp xỉ của chuỗi con $P(p_i, \dots, p_j)$ chỉ đơn giản là một đường thẳng kết nối điểm dữ liệu p_i, p_j .

Thuật toán *PLR* tạo một xấp xỉ của chuỗi thời gian, vì vậy $m/2$ phân đoạn được sử dụng để xấp xỉ với m là chiều dài chuỗi thời gian và lặp lại phép gộp sao cho chi phí của phân đoạn là thấp nhất. Thực hiện tạo xấp xỉ như vậy cho đến khi đủ số phân đoạn theo yêu cầu thì dừng lại. Khi phân đoạn mới được tạo ra bằng cách gộp phân đoạn S_i và S_{i+1} lại với nhau thì giá trị của phân đoạn mới gộp với phân đoạn bên phải của nó $S_i + 1$ phải được tính toán.

Phương pháp *PIP* (*Perceptually Important Points*) thực hiện giảm số chiều bằng cách giữ lại các điểm nổi bật. Phương pháp này thích hợp để sử dụng cho các mẫu phân tích trong các ứng dụng tài chính. Cho chuỗi thời gian P với n điểm dữ liệu P_1, P_2, \dots, P_n . Tất cả các điểm dữ liệu trong P có thể được sắp xếp dựa vào độ quan trọng của nó thông qua quá trình xử lý định danh *PIP*. Điểm dữ liệu thứ nhất P_1 trong chuỗi dữ liệu tương ứng với *PIP* đầu tiên và điểm dữ liệu cuối cùng P_n tương ứng với *PIP* thứ hai. *PIP* tiếp theo được tìm thấy là một điểm trong P có khoảng cách lớn nhất so với hai *PIP* đầu tiên. *PIP* thứ tư cũng là một điểm trong P có khoảng cách theo chiều dọc lớn nhất đến đường thẳng nối hai *PIP* liên kề, giữa *PIP* đầu tiên và thứ hai hoặc là giữa *PIP* thứ hai và *PIP* cuối cùng. Xử lý *PIP* tại vị trí tiếp theo cho đến khi tất cả các điểm trong P được đưa vào một danh sách có thứ tự L hoặc là đủ số lượng *PIP* được yêu cầu. *PIP* thứ bảy được xác định trong ví dụ về chuỗi thời gian tại **Hình 2.3**.



Hình 2.3. Giảm chuỗi thời gian bằng cách giữ lại các điểm quan trọng [8]

Một số phương pháp biểu diễn khác như phương pháp CPM (*Critical Point Model*) và phương pháp biểu diễn cấp độ cao dựa trên một chuỗi các điểm tới hạn được đề xuất để phân tích dữ liệu tài chính. Phương pháp SAX (*Symbolic Aggregate Approximation*) thực hiện chuyển đổi kết quả thu được từ phương pháp PAA thành chuỗi ký tự. Phương pháp PCA (*Principal Component Analysis*) là một kỹ thuật đa chiều phổ biến dùng để xử lý việc thống kê đa chiều bằng phương pháp định lượng và nó được sử dụng để phân tích dữ liệu tài chính.

2.3. Các kỹ thuật khai thác dự báo dữ liệu chuỗi thời gian

2.3.1. Phân lớp

Phân lớp là kỹ thuật quen thuộc và phổ biến nhất trong các kỹ thuật khai thác dữ liệu. Phân lớp được ứng dụng để nhận biết hình ảnh và mẫu, lọc thư rác, chẩn đoán y tế, phát hiện các sự cố trong ứng dụng công nghiệp. Phân lớp dữ liệu đầu vào thành các nhóm đã được định nghĩa trước đó. Hình thức này được gọi là học có giám sát bởi vì các lớp được xác định trước nhằm kiểm tra dữ liệu. Bộ dữ liệu được xây dựng sẵn từ trước sẽ được dùng trong quá trình huấn luyện và học nhằm nhận biết mẫu đáng quan tâm nhất. Nhận dạng mẫu là một loại phân lớp với các mẫu đầu vào được phân lớp thành một trong vài lớp dựa trên sự tương đồng của nó với các lớp đã được xác định trước đó. Hai phương pháp phổ biến nhất trong phân lớp chuỗi thời gian là phân loại Nearest Neighbor và cây quyết định. Phương pháp Nearest Neighbor cung cấp độ đo tương đồng với các đối tượng được phân lớp để quyết định phân lớp tốt nhất dựa trên các dữ liệu hiện có đã được phân lớp. Đối với cây quyết định, một tập các luật được suy ra từ dữ liệu huấn luyện. Tập luật này được áp dụng với bất kỳ bộ dữ liệu mới nào được phân lớp. Cây quyết định được định nghĩa cho dữ liệu thực tế nhưng việc cố gắng áp dụng dữ liệu chuỗi thời gian có thể là sai lầm vì chiều cao và mức độ nhiễu sẽ ảnh hưởng đến độ sâu của cây, làm cho cây dày đặc dữ liệu. Do đó, một số nhà nghiên cứu đề nghị sử dụng cây hồi quy thay cho cây quyết định.

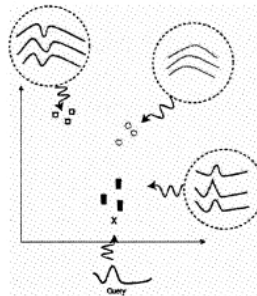
2.3.2. Chỉ mục

Truy vấn theo nội dung là việc so khớp chuỗi. Việc so khớp chuỗi bao gồm so khớp toàn bộ và so khớp chuỗi con.

+ So khớp toàn bộ: Một chuỗi thời gian truy vấn được so khớp với cơ sở dữ liệu chuỗi thời gian riêng biệt để xác định cái giống với phần cần truy vấn.

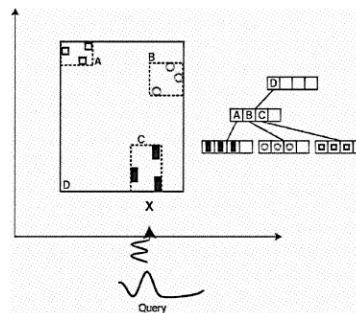
+ So khớp chuỗi con: Một chuỗi thời gian truy vấn ngắn được so khớp với chuỗi thời gian dài hơn bằng cách trượt nó theo chuỗi dài hơn để tìm kiếm vị trí phù hợp nhất.

Cấu trúc chỉ mục dạng vector: Cấu trúc chỉ mục này có số chiều giống với số chiều của dữ liệu sau khi được nén. Các chuỗi thời gian ban đầu được nén lại bằng phương pháp giảm số chiều và các vector đa chiều kết quả có thể được nhóm thành các cụm tương tự bằng cách sử dụng nhiều kỹ thuật tạo chỉ mục dạng vector.



Hình 2.4. Giảm số chiều của dữ liệu chuỗi thời gian thành 2 chiều [14]

Cấu trúc chỉ mục có thể được sử dụng ở hai dạng khác nhau gồm phân cấp hoặc không phân cấp. Chỉ mục dựa trên vector phân cấp phổ biến nhất là cây *R-tree*. Các chỉ mục được mô tả tại các nút lá của cây và quyết định hình dạng của cấu trúc cây bằng cách sử dụng các siêu hình chữ nhật chồng lên nhau (**Hình 2.5**).



Hình 2.5. Cấu trúc chỉ mục phân cấp sử dụng cây R-tree [14]

Cấu trúc chỉ mục dựa trên khoảng cách: Cấu trúc chỉ mục này được sử dụng linh hoạt hơn cấu trúc chỉ mục dạng vector thậm chí vẫn linh hoạt hơn đối với các chỉ mục có số chiều cao (có thể từ 20 đến 30 chiều) vì chỉ cần khoảng cách giữa hai đối

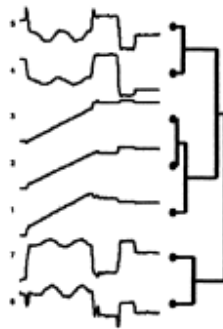
tượng để tạo thành các chỉ mục này. Vì vậy, cấu trúc chỉ mục dựa trên khoảng cách không thực hiện gom cụm bằng đặc trưng mà thông qua khoảng cách giữa các đối tượng liên quan.

2.3.3. Gom cụm

Gom cụm tương tự như phân lớp dữ liệu thành các nhóm. Tuy nhiên, các nhóm này không được xác định trước mà xác định bằng chính dữ liệu đó dựa trên độ tương tự giữa hai chuỗi thời gian. Đây là hình thức học không có giám sát.

Gom cụm thường dựa vào độ tương tự giữa các thuộc tính đã được định nghĩa trước. Có 2 phương pháp gom cụm chuỗi thời gian là gom cụm phân cấp và gom cụm không phân cấp.

+ Gom cụm phân cấp tính khoảng cách giữa các cặp dữ liệu, và sau đó gộp các cụm giống nhau theo thứ tự từ dưới lên mà không cần cung cấp số lượng cụm. Đây là một trong những công cụ tốt nhất để đánh giá dữ liệu bằng cách tạo sơ đồ cây của một vài chuỗi thời gian từ miền dữ liệu hữu ích, được thể hiện trong **Hình 2.6**. Tuy nhiên, gom cụm chỉ áp dụng cho các bộ dữ liệu nhỏ do độ phức tạp tính toán lớn.



Hình 2.6. Gom cụm phân cấp của chuỗi thời gian [14]

+ Gom cụm không phân cấp sử dụng thuật toán *K-means* để tối ưu hóa hàm mục tiêu bằng giảm thiểu tổng các lỗi bên trong cụm. Số lượng các cụm là K phải được xác định trước.

Gom cụm được sử dụng trong nhiều lĩnh vực ứng dụng như sinh học, y học, nhân chủng học, thị trường và kinh tế. Có 2 loại gom cụm chuỗi thời gian: gom cụm toàn bộ và gom cụm chuỗi con.

+ Gom cụm toàn bộ là phân nhóm các đối tượng rời rạc. Cho bộ dữ liệu thời gian riêng biệt, các chuỗi thời gian giống nhau sẽ vào cùng một nhóm.

+ Gom cụm chuỗi con thực hiện trên từng chuỗi thời gian con riêng biệt được trích xuất từ các chuỗi thời gian dài ứng với một cửa sổ trượt. Gom cụm chuỗi con là một bước tiền xử lý chung cho nhiều thuật toán khai thác mẫu nhằm tìm ra các luật chuỗi thời gian.

2.3.4. Dự báo

Dự báo có thể xem như là một loại của gom cụm hoặc phân lớp, chỉ khác ở chỗ là dự báo một trạng thái trong tương lai chứ không phải ở hiện tại. Dự báo được ứng dụng nhằm đưa ra các dự báo về thời tiết (thiên tai, lũ lụt, bão...), dịch bệnh, rủi ro chứng khoán,... Nhiều ứng dụng dự báo chuỗi thời gian có thể thấy trong các lĩnh vực kinh tế với thuật toán dự báo điển hình liên quan đến phân tích hồi quy. Dự báo kết quả tương lai dựa trên các xu hướng và thống kê đã có. Nhiều kỹ thuật được đề xuất nhằm tăng độ chính xác chuỗi thời gian dự báo như sử dụng kỹ thuật mạng nơron, các kỹ thuật giảm thứ nguyên.

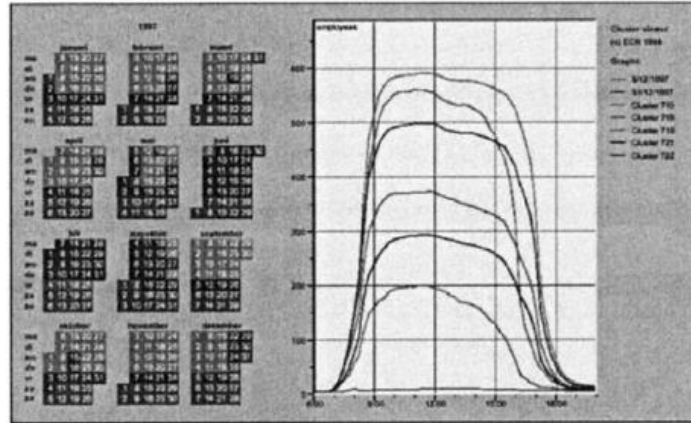
2.3.5. Tổng hợp

Dữ liệu chuỗi thời gian có thể dài và khổng lồ. Vì vậy, tổng hợp dữ liệu có thể hữu ích và thật sự cần thiết. Có thể sử dụng ngôn ngữ tự nhiên, hình ảnh hóa hoặc tóm tắt đồ họa để trích xuất thông tin hữu ích hoặc có ý nghĩa từ dữ liệu. Phát hiện bất thường và khám phá mô-tip là trường hợp đặc biệt của tổng hợp, chỉ các mẫu bất thường được quan tâm và báo cáo. Tổng hợp cũng có thể xem như là một trường hợp đặc biệt của gom cụm mà các bản đồ dữ liệu được phân thành các tập con dựa vào mô tả văn bản hoặc hình ảnh, cung cấp một cái nhìn ở mức độ cao hơn về dữ liệu. Điều này mô tả sự đơn giản hóa của dữ liệu. Tổng hợp có thể được thực hiện ở nhiều độ chi tiết và nhiều kích thước khác nhau.

Một vài tiếp cận phổ biến để trừu tượng hóa bộ dữ liệu chuỗi thời gian khổng lồ là *TimeSearcher*, *Cluster and Calendar-Based Visualization*, *Spiral* và *VizTree*.

+ *TimeSearcher* là chuỗi thời gian truy vấn có tính chất thăm dò, công cụ hình ảnh hóa cho phép người sử dụng khôi phục chuỗi thời gian bằng cách tạo ra các

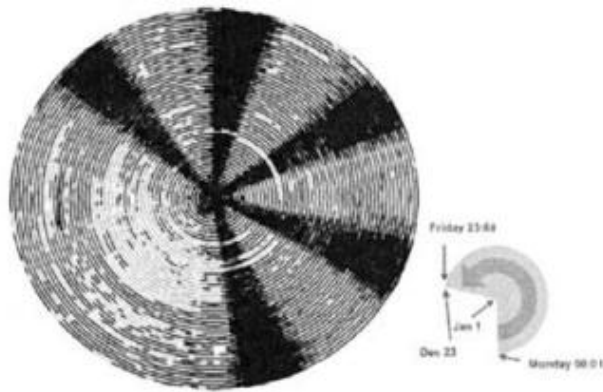
truy vấn. Tuy nhiên, một vài tri thức về các bộ dữ liệu có thể cần thiết trước và người dùng cần phải có một ý tưởng chung về những gì tìm kiếm và quan tâm.



Hình 2.7. Một ví dụ về trừu tượng hóa lịch và gom cụm [14]

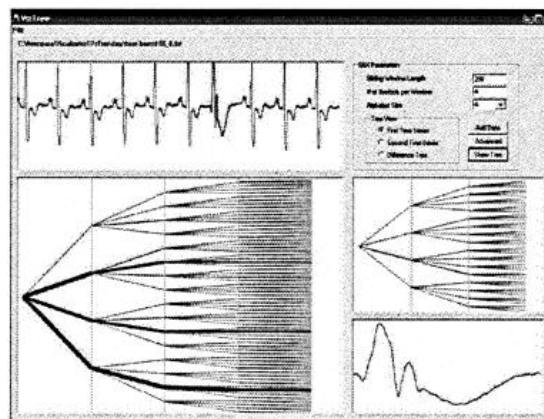
+ *Cluster and Calendar-Based Visualization* là một hệ thống trừu tượng mà khối dữ liệu chuỗi thời gian được phân thành chuỗi các mẫu theo ngày. Các mẫu theo ngày đó được nhóm lại bằng thuật toán gom cụm từ dưới lên. Hệ thống hiển thị các mẫu đại diện bởi trung bình cụm cùng với một lịch theo từng mã màu mà nó thuộc về. Từ các mẫu đó có thể khám phá ra luật đầy tiềm năng. **Hình 2.7** minh họa một khung nhìn tổng quát về trừu tượng hóa lịch và gom cụm. Các mẫu được tìm thấy được liên kết với một lịch, từ đó có thể khám phá ra các luật đơn giản như “Trong những tháng mùa đông thì lượng điện tiêu thụ nhiều hơn những tháng mùa hè”.

+ *Spiral* vẽ định kỳ từng phần của chuỗi thời gian lên trên một vòng tròn và các thuộc tính như màu, độ dày được sử dụng để mô tả các giá trị của dữ liệu. Tiếp cận chính của kỹ thuật này là xác định các cấu trúc định kỳ trong dữ liệu. Tuy nhiên, cách tiếp cận này có hạn chế đối với chuỗi thời gian nếu chuỗi thời gian không thể hiện theo định kỳ hoặc thời gian không rõ ràng. **Hình 2.8** minh họa việc sử dụng điện năng hàng năm.



Hình 2.8. Ví dụ về sử dụng điện năng hàng năm [14]

+ *VizTree* được sử dụng nhằm khám phá ra các mẫu dữ liệu không được biết hoặc ít được biết trước đó. Nó cung cấp một tóm tắt tổng thể và có khả năng tìm thấy cấu trúc tiềm năng bị ẩn dấu bên trong dữ liệu. Cách tiếp cận này đầu tiên chuyển đổi các chuỗi thời gian thành một đại diện tượng trưng và mã hóa dữ liệu thành cây hậu tố với tần số và các tính chất khác của mẫu được ánh xạ thành các màu và các tính chất tượng trưng khác. Dữ liệu chuỗi thời gian gốc là liên tục nhưng cấu trúc cây cần dữ liệu rời rạc. Trong cây *VizTree*, các mẫu phổ biến có thể phát hiện bởi các nhánh chứa dữ liệu dày đặc, các mẫu bất thường đơn giản có thể tìm thấy trong các nhánh chứa dữ liệu thưa thớt. **Hình 2.9** thể hiện một cây con vừa được dùng để phát hiện đặc tính vừa được dùng để phát hiện bất thường trên dữ liệu ECG.

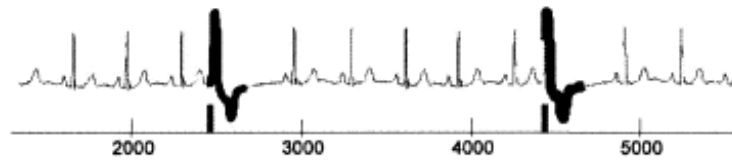


Hình 2.9. Minh họa các bất thường trên dữ liệu ECG [14]

2.3.6. Phát hiện bất thường

Trong khai thác và giám sát dữ liệu chuỗi thời gian, vấn đề phát hiện các mẫu bất thường thu hút được nhiều sự chú ý. Ngược lại với so khớp chuỗi con, phát hiện

bất thường xác định các mẫu không được biết trước đó. Đây là vấn đề đặc biệt khó khăn bởi vì những gì tạo nên bất thường rất khác nhau tùy thuộc vào công việc thực hiện. Xác định bất thường bằng cách căn cứ vào tần số xuất hiện, nếu tần số xuất hiện khác hẳn so với dự kiến thì đó là bất thường (cho phép nhìn thấy trước dữ liệu). Vấn đề phát hiện bất thường trong chuỗi thời gian được khái quát hóa thành phát hiện các mẫu hữu ích (không nhất thiết phải bất thường). Phát hiện bất thường liên quan chặt chẽ với tổng hợp.



Hình 2.10. Một phát hiện bất thường từ dữ liệu kiểm thử MIT-BIH [14]

2.3.7. Phân đoạn

Phân đoạn chuỗi thời gian thường được gọi là thuật toán giảm chiều. Các phân đoạn được tạo ra có thể là đa thức với bậc bất kỳ. Biểu diễn phổ biến nhất của các phân đoạn là các hàm tuyến tính. Biểu diễn tuyến tính từng phần PLR (*Piecewise Linear Representation*) được minh họa trong **Hình 2.11**.



Hình 2.11. Một phân đoạn chuỗi thời gian [14]

Các thuật toán phân đoạn chuỗi thời gian được chia thành ba loại:

- + Cửa sổ trượt (*Sliding-Windows*): Một phân đoạn được phát sinh cho đến khi nó vượt quá một số lỗi bị ràng buộc. Việc lặp đi lặp lại với điểm dữ liệu kế tiếp không bao gồm phân đoạn mới xấp xỉ.
- + *Top-Down*: Chuỗi thời gian được phân chia theo phương pháp đệ quy cho đến khi một vài điều kiện dừng được đáp ứng.

+ *Bottom-Up*: Bắt đầu xấp xỉ từ cái tốt nhất, các phân đoạn được gộp lại cho đến khi một vài điều kiện dừng được đáp ứng.

Có nhiều cách để xác định chất lượng của thuật toán phân chia. Rõ ràng nhất là thực hiện đo các lỗi tái thiết cho một số phân đoạn cố định. Các lỗi tái thiết chỉ đơn giản là khoảng cách O-clit giữa dữ liệu gốc và đại diện phân đoạn.

2.4. Phân tích và chọn lựa các kỹ thuật

Các kỹ thuật phân lớp, chỉ mục, gom cụm, tổng hợp, phát hiện bất thường, phân đoạn chỉ xử lý dữ liệu nhằm tìm ra các mẫu quan tâm nhất, tìm ra tập các luật. Trong khi đó, luận văn đề xuất khai thác dữ liệu chuỗi thời gian nhằm đưa ra các dự báo dữ liệu tương lai, cụ thể là dự báo giá cổ phiếu. Do đó, với các kỹ thuật khai thác dữ liệu chuỗi thời gian đã được trình bày thì chỉ có kỹ thuật dự báo dựa trên mạng nơron là phù hợp với yêu cầu của luận văn.

2.5. Tổng kết chương

Chương này đã giới thiệu một cách tổng quát nhất về dữ liệu chuỗi thời gian. Trình bày các lý thuyết cơ bản nhất về dữ liệu chuỗi thời gian như loại dữ liệu nào thì được gọi là dữ liệu chuỗi thời gian, một số cách biểu diễn dữ liệu chuỗi thời gian, các ứng dụng của dữ liệu chuỗi thời gian. Khai thác dữ liệu chuỗi thời gian với các kỹ thuật như phân lớp, chỉ mục, gom cụm, dự báo, tổng kết, phát hiện bất thường và phân đoạn. Sau khi phân tích và chọn lựa các kỹ thuật thì chỉ có kỹ thuật dự báo là phù hợp với yêu cầu luận văn. Kỹ thuật dự báo sử dụng mạng nơron để đưa ra kết quả dự báo cho tương lai. Từ các cơ sở lý thuyết đã tiếp cận này, luận văn khai thác dữ liệu cổ phiếu tại Cophieu68 nhằm đưa ra dự báo giá cổ phiếu cho tương lai. Vì vậy, ngoài việc sử dụng kỹ thuật mạng nơron, luận văn kết hợp thêm hệ suy luận mờ để dự báo giá cổ phiếu. Các nội dung này sẽ lần lượt được tiếp cận trong các chương sau.

CHƯƠNG 3

CÁC LÝ THUYẾT CƠ SỞ XÂY DỰNG HỆ SUY LUẬN ANFIS

3.1. Giới thiệu

Nhằm xây dựng mạng neuro-fuzzy ANFIS để dự báo giá cổ phiếu, trong chương này lần lượt tiếp cận hai đối tượng nghiên cứu chính đó là hệ suy luận mờ và mạng nơron nhân tạo ANN. Đối với hệ suy luận mờ, nội dung của chương sẽ trình bày về các tham số mờ, các tập mờ, các phương pháp mờ hóa, khử mờ để từ đó ứng dụng phương pháp nội suy mờ. Đối với mạng nơron nhân tạo ANN, nội dung của chương trình bày về tổng quan mạng nơron nhân tạo, phương pháp huấn luyện mạng nơron nhân tạo để từ đó thực hiện quá trình kiểm thử của mạng nơron nhân tạo. Từ hai đối tượng nghiên cứu cơ sở trên, luận văn định hướng xây dựng hệ dự báo giá cổ phiếu dựa trên mạng neuro-fuzzy ANFIS.

3.2. Hệ suy luận mờ

3.2.1. Tổng quan về suy luận mờ

Việc áp dụng tính toán mềm (*soft computing*) và các hệ điều khiển mờ (*fuzzy control system*) ngày càng được phổ biến và áp dụng nhiều hơn trong cuộc sống hàng ngày bởi tính ứng dụng mạnh mẽ của chúng. Các khái niệm về hệ thống thông minh, hệ thống tự động cũng không còn xa lạ đối với cuộc sống hàng ngày, các chương trình tự học, tự suy nghĩ giống như con người ngày càng được hoàn thiện hơn về mặt lý thuyết cũng như về mặt ứng dụng của chúng [28].

Khái niệm mờ (*vague, fuzzy concept*) cũng đã dần nhiều trong cuộc sống hàng ngày của con người. Việc mô tả các hiện tượng, mô tả các cảm xúc đều là những biểu hiện của khái niệm mờ.

Mô hình mẫu cho kỹ thuật tính toán mềm là tư duy của con người, giải quyết các vấn đề không chắc chắn và không chính xác dựa trên những phương pháp tính toán và lập luận logic truyền thống.

Trong logic nhị phân, tập giá trị bao gồm $\{true, false\}$ hoặc tương ứng $\{1, 0\}$. Việc tiếp cận bằng logic nhị phân (logic kinh điển) đã không còn phù hợp với việc

biểu diễn tri thức của con người vì bản chất và suy nghĩ của con người vốn trừu tượng và ngôn ngữ sử dụng có độ chính xác không rõ ràng [32].

Để tiếp cận ngày càng gần hơn tư duy của con người cần phải mở rộng tập các giá trị chân lý, từ đó tạo thành các logic 3 giá trị, logic đa trị, logic mờ, logic xác suất, logic khoảng,...

Những ý tưởng cơ bản về những lập luận thiếu chính xác đã được đề cập bởi các nhà logic học và nhà triết học Max Black trong những năm cuối thế kỷ XIX và đầu thế kỷ XX.

Mô hình toán học đầu tiên của các khái niệm mờ được Lotfi.A.Zadeh đề xuất vào năm 1965 dựa trên khái niệm tập mờ. Sau đó, Zadeh đã đề xuất và phát triển lý thuyết lập luận xấp xỉ dựa trên khái niệm biến ngôn ngữ (*linguistic variable*) và logic mờ (*fuzzy logic*) [32].

Các giá trị của một biến ngôn ngữ được xây dựng từ các phần tử sinh nguyên thủy (*primary term*) của các biến đó như *very*, *more or less*,... và các liên từ (*connectives*) ví dụ như: *AND*, *OR*,... [28]. Theo Zadeh, logic mờ làm cơ sở cho phương pháp lập luận xấp xỉ là logic giá trị ngôn ngữ.

Quá trình lập luận xấp xỉ là quá trình tìm ra các kết luận không chắc chắn bằng phương pháp suy diễn theo nghĩa xấp xỉ từ một họ các tiền đề không chắc chắn, lập luận không chắc chắn nằm ngoài khả năng của logic kinh điển.

Lý thuyết lập luận xấp xỉ đã có nhiều ứng dụng khác nhau, đặc biệt là trong các hệ chuyên gia mờ (*fuzzy expert system*), điều khiển dựa trên logic mờ (*fuzzy logic control*)...

Trong khuôn khổ luận văn tìm hiểu các lý thuyết về logic mờ, các dạng tập mờ, tìm hiểu cách biểu diễn tập giá trị chân lý ngôn ngữ cho tập mờ. Tìm hiểu mối quan hệ giữa các dạng biểu diễn tập mờ.

3.2.2. Logic mờ

Logic mờ là một hệ logic tổng quát hoá dạng logic hai giá trị cổ điển cho việc lập luận dưới góc độ không chắc chắn. Đây là một hệ thống tính toán và suy luận gần

đúng dựa vào một tập các lý thuyết và kỹ thuật để xây dựng nên các tập mờ là những lớp của đối tượng không có đường bao hình dạng [32].

Hai đặc trưng chính của logic mờ:

+ Một mô hình toán thể hiện tri thức của con người liên quan đến các khái niệm mập mờ.

+ Một cơ chế tự nhiên và hiệu quả cho các giải pháp có tính hệ thống của các vấn đề phức tạp được mô tả bởi thông tin không chắc chắn và thiếu chính xác.

Logic mờ khá phù hợp trong việc điều khiển một quá trình hoặc một hệ thống có tính phi tuyến hoặc quá phức tạp khi sử dụng các phương pháp điều khiển truyền thống.

3.2.3. Tập mờ và biến ngôn ngữ

3.2.3.1. Tập mờ

Tập mờ là công cụ cho phép thể hiện các thuật ngữ trong ngôn ngữ tự nhiên mang tính mập mờ theo toán học một cách cụ thể. Tập mờ là một sự tổng quát hoá của một lớp hay một tập rõ nghĩa là một tập mờ có thể chứa một phần các phần tử của nó. Ngược lại, phần tử x của một tập rõ có thể thuộc hoặc không thuộc tập đó.

Hàm đặc tính $X_A : U \rightarrow \{0,1\}$ của một tập rõ A ánh xạ các phần tử x của một tập nền U thành 0 hoặc 1.

L.A.Zadeh đã đưa ra khái niệm giá trị chân lý ngôn ngữ **TRUTH** = $\{true, not\ true, very\ true, more\ or\ less\ true, very\ very\ true, essentially\ true, very\ not\ true, not\ very\ true, \dots false, not\ false, very\ false, \dots, not\ very\ true\ and\ not\ very\ false\}$ là cơ sở cho logic mờ [28].

Để “**tính toán**” các giá trị chân lý này, tiến hành “**diễn dịch**” chúng thành các tập mờ trên không gian tham chiếu $[0, 1]$.

Định nghĩa: Cho U là một tập khác rỗng (tập vũ trụ). Một tập mờ A trên U là tập có thứ tự như sau:

$$A = \{(\mu(u), u) \mid \mu : U \rightarrow [0,1], u \in U\} \quad (3.1)$$

Trong đó: $\mu(u)$ là mức độ thuộc của phần tử u trong A .

μ là hàm thuộc của tập mờ A .

Trong trường hợp U hữu hạn, $U = \{u_1, u_2, \dots, u_n\}$ thì tập mờ A kí hiệu như sau:

$$A = \sum_{i=1}^n \mu(u_i) / u_i \quad (3.2)$$

Trong trường hợp U liên tục thì tập mờ A được kí hiệu như sau:

$$A = \int_{u \in U} \mu(u) / u \quad (3.3)$$

Nhận xét:

- + Mỗi tập mờ trên U như một hàm từ U vào đoạn $[0, 1]$.
- + Khái niệm tập mờ là mở rộng của khái niệm tập con của một tập cho trước.
- + Hàm thuộc là một hàm tổng quát của hàm đặc trưng.
- + Các lý thuyết tập hợp: *hội, giao, lấy phần bù* cũng được mở rộng cho các tập mờ.

3.2.3.2. Biến ngôn ngữ

Biến ngôn ngữ (*linguistic variable*) là các biến mà có các giá trị là các giá trị ngôn ngữ. Các giá trị này được xây dựng từ các phần tử sinh nguyên thủy (*primary term*) của biến đó.

Biến ngôn ngữ đề cập đến những từ cũng như các giá trị của nó có tính chất biến đổi trong ngôn ngữ tự nhiên. Nói cách khác, biến ngôn ngữ là các thuật ngữ ngôn ngữ, không lấy giá trị số và được mô tả một cách mập mờ bằng các từ. Các từ như “*rất nhỏ*”, “*nhỏ*”, “*vừa*”, ... là những ví dụ về biến ngôn ngữ được sử dụng trong đời sống hằng ngày. Những giá trị của biến ngôn ngữ đưa đến khái niệm về tập mờ.

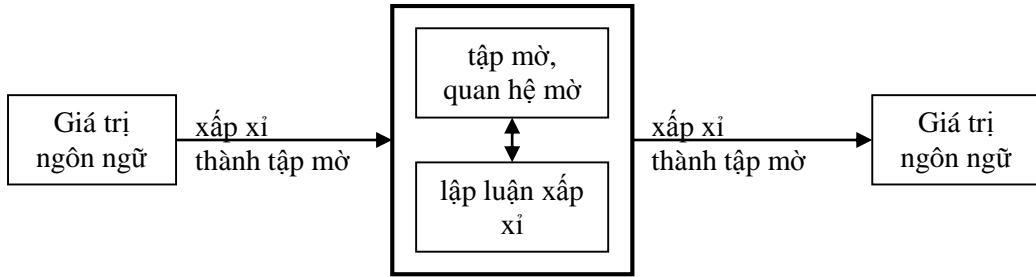
Biến ngôn ngữ X là một ánh xạ từ không gian nền U vào một không gian con R_x của tập số thực R . $U : X \rightarrow R_x \subset R$

Các giá trị ngôn ngữ của một biến ngôn ngữ được biến đổi thành các tập mờ trên một không gian tham chiếu cụ thể tùy thuộc vào ứng dụng. Phép biến đổi này được thực hiện từng bước như sau: [32]

- + Biến đổi các giá trị ngôn ngữ nguyên thủy thành các tập mờ.

+ Tập mờ tương ứng với giá trị ngôn ngữ hợp thành được tính bằng cách áp dụng các toán tử cho các tập mờ của giá trị ngôn ngữ nguyên thủy.

Việc biến đổi này được thực hiện dựa trên việc phân tích ngữ nghĩa của các giá trị ngôn ngữ. Trong các ứng dụng của hệ mờ (*fuzzy system*) các giá trị ngôn ngữ thường được sử dụng để mô tả tri thức chuyên gia về các thuộc tính định lượng của hệ thống.



Hình 3.1. Mô hình lập luận xấp xỉ mờ trên giá trị ngôn ngữ

3.2.4. Biểu diễn tập mờ

3.2.4.1. Tập mờ hàm số mũ

Theo L.A. Zadeh đề xuất, cho các trạng từ nhấn *very* hay *more or less*. Từ đó, tạo ra một họ các tập mờ là hàm mũ từ hai tập mờ sinh tương ứng *true*, và *false*

Ví dụ: $very\ true = true^2$; $very\ very\ true = true^4$; $absolutely\ true = true^\infty$

$more\ or\ less\ true = true^{0.5}$; $more\ or\ less\ very\ true = true^{1.5}$

$very\ more\ or\ less\ true = true^{0.75}$

Một cách tổng quát, cho giá trị chân lý $\sigma true$, sẽ có thể tính toán được một số mũ $\beta = exponent(\sigma)$ tương ứng, sao cho $\sigma true = true^\beta$. Tương tự như vậy, cũng có $\sigma false = false^\beta$.

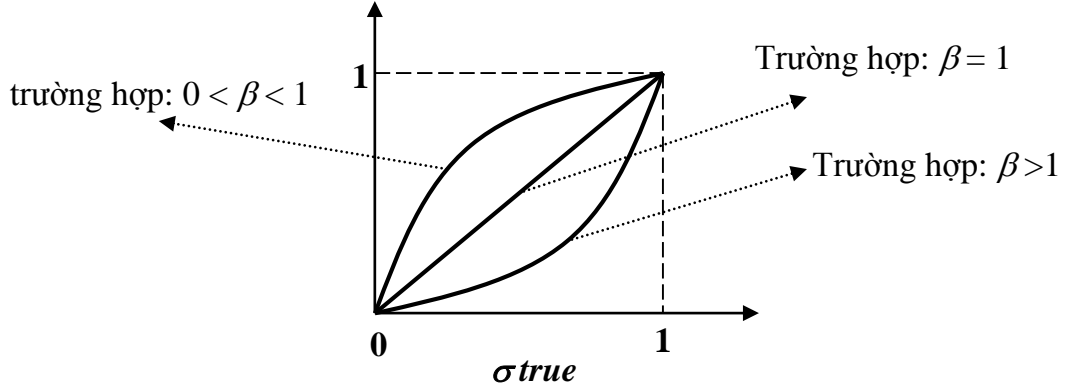
Các tập mờ sinh *true*, *false* được lựa chọn:

$$true = \int_{[0,1]} t/t \text{ và } false = \int_{[0,1]} (1-t)/t. \quad (3.4)$$

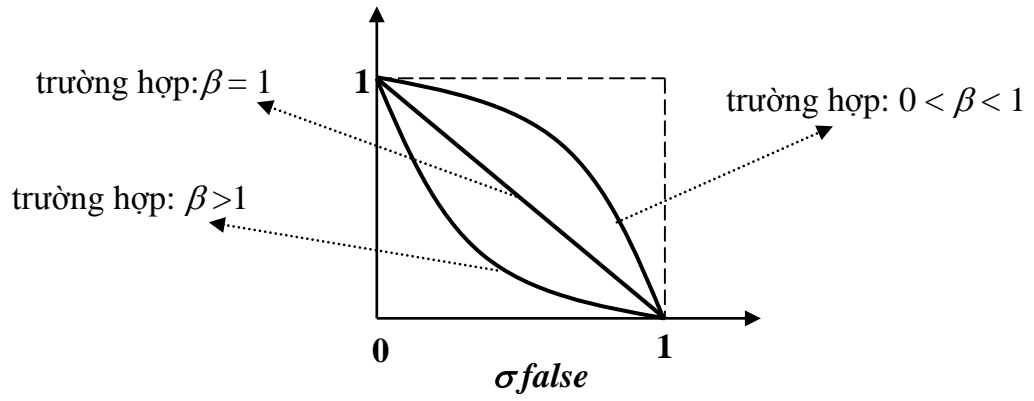
Khi đó :

$$\sigma true = \int_{[0,1]} t^\beta / t \quad \sigma false = \int_{[0,1]} (1-t)^\beta / t \quad (3.5)$$

Khi σ tăng ngữ nghĩa của *true* thì $\beta > 1$, khi σ làm giảm ngữ nghĩa của *true* thì $0 < \beta < 1$, khi $\sigma = 0$ thì $\beta = 1$. Tương tự như vậy cho *false*.



Hình 3.2. Biểu diễn tập mờ hàm mũ ứng với giá trị chân lý TRUE



Hình 3.3. Biểu diễn tập mờ hàm mũ ứng với giá trị chân lý FALSE

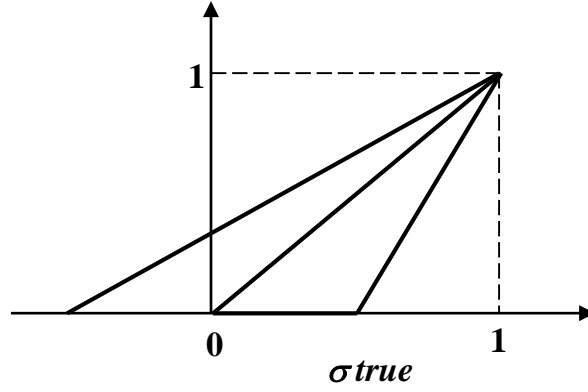
3.2.4.2. Tập mờ tam giác

Các tập mờ sinh *true*, *false* được lựa chọn

$$true = \int_{[0,1]} t/t \text{ và } false = \int_{[0,1]} (1-t)/t \quad (3.6)$$

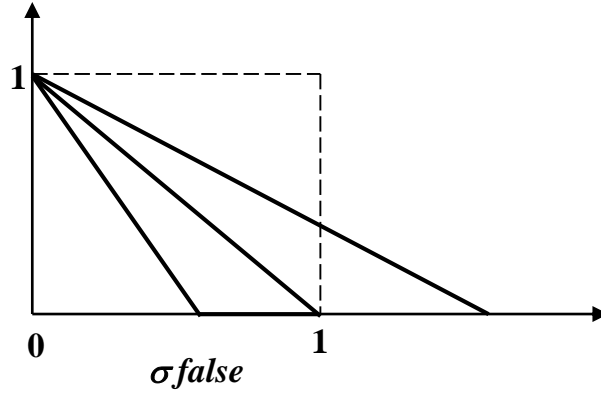
σ_{true} và σ_{false} được xác định như sau:

$$\sigma_{true} = \begin{cases} \int_{[0,a]} 0/t + \int_{[a,1]} \frac{t-a}{1-a} & \text{khi } a \geq 0 \\ \int_{[0,1]} \frac{t-a}{1-a} / t & \text{khi } a \leq 0 \end{cases} \quad (3.7)$$



Hình 3.4. Biểu diễn tập mờ tam giác ứng với giá trị chân lý TRUE

$$\sigma_{false} = \begin{cases} \int_{[0,a]} 1 - \frac{t}{a} / t + \int_{[a,1]} 0 / t & \text{khi } a \leq 1 \\ \int_{[0,1]} 1 - \frac{t}{a} / t & \text{khi } a \geq 1 \end{cases} \quad (3.8)$$



Hình 3.5. Biểu diễn tập mờ tam giác ứng với giá trị chân lý FALSE

Tham số a phụ thuộc vào σ .

Khi σ làm tăng ngữ nghĩa của *true* thì $0 < a \leq 1$, khi làm giảm thì $a < 0$, khi $\sigma = 0$ thì $a = 0$.

Tương ứng với *false* là $0 \leq a < 1$, $a > 1$ và $a = 1$.

3.2.4.3. Tập mờ tuyến tính

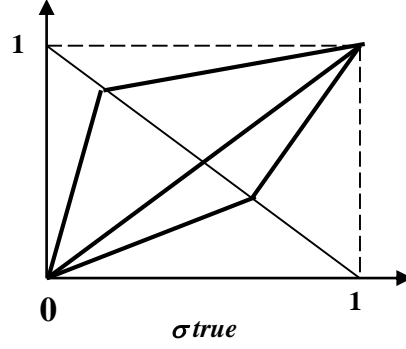
Nhằm tuyến tính hoá hàm số mũ với các tập mờ sinh *true*, *false* được lựa chọn [32].

$$true = \int_{[0,1]} t / t \text{ và } false = \int_{[0,1]} (1-t) / t \quad (3.9)$$

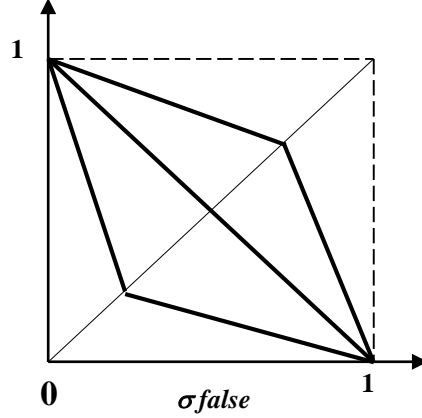
σ_{true} và σ_{false} được xác định như sau:

$$\sigma_{true} = \int_{[0,b]} \frac{1-b}{b} t/t + \int_{[b,1]} \frac{b}{1-b} t + \frac{1-2b}{1-b} /t \quad (3.10)$$

$$\sigma_{false} = \int_{[0,b]} \frac{b-1}{b} t + 1/t + \int_{[b,1]} \frac{b}{b-1} t - \frac{b}{b-1} /t \quad (3.11)$$



Hình 3.6. Biểu diễn tập mờ tuyến tính với giá trị chân lý TRUE



Hình 3.7. Biểu diễn tập mờ tuyến tính với giá trị chân lý FALSE

Tham số của họ tập mờ này là $b \in [0, 1]$, với σ_{true} , khi tăng ngữ nghĩa của *true* thì $b > 0.5$, nếu làm giảm ngữ nghĩa thì $b < 0.5$ và nếu $\sigma = 0$ thì $b = 0.5$. Tương tự cho trường hợp σ_{false}

3.2.5. Các phương pháp khử mờ

3.2.5.1. Phương pháp khử mờ trung bình tổng quát

Khi áp dụng phương pháp khử mờ sẽ cho ánh xạ định lượng ngữ nghĩa. Phương pháp khử mờ trung bình tổng quát với tham số mũ $\gamma > 0$ có công thức được mô tả như sau:

Cho tập mờ rời rạc $A = \sum_{t \in U} \frac{\mu_A(t)}{t}$, khi đó, kết quả khử mờ sẽ là:

$$t^* = \frac{\sum_{t \in U} \mu_A(t)^\gamma t}{\sum_{t \in U} \mu_A(t)^\gamma} \quad (3.12)$$

Trong đó: $\mu_A(t)$ là hàm thuộc của phần tử t trong tập mờ A , t^* là giá trị khử mờ và γ là tham số khử mờ ($\gamma > 0$). Sau khi thực hiện khử mờ sẽ có được giá trị rõ.

Ví dụ:

Cho tập mờ liên tục $A = \int_{[0,1]} \mu_A(t)/t$, thì kết quả khử mờ sẽ là:

$$t^* = \frac{\int_0^1 [(\mu_A(t))^\gamma t] dt}{\int_0^1 [(\mu_A(t))^\gamma] dt} \quad (3.13)$$

Theo như trên, $\gamma = 1$ thì đó là *phương pháp khử mờ trung bình*, khi $\gamma \rightarrow \infty$ thì đó là *phương pháp khử mờ lấy giá trị có độ thuộc cực đại*.

Tham số γ được sử dụng để thực hiện điều khiển sao cho khi khử mờ các tập mờ sẽ thu được đúng giá trị định lượng ngữ nghĩa đã được định nghĩa trước.

3.2.5.2. Phương pháp khử mờ hàm số mũ

Với σ_{true} :

$$t^* = \frac{\int_0^1 [(\mu_{\sigma_{true}}(t))^\gamma t] dt}{\int_0^1 (\mu_{\sigma_{true}}(t))^\gamma dt} = \frac{\int_0^1 [(t^\beta)^\gamma t] dt}{\int_0^1 (t^\beta)^\gamma dt} = \frac{\int_0^1 [t^{\beta\gamma+1}] dt}{\int_0^1 t^{\beta\gamma} dt} = \frac{\beta\gamma + 1}{\beta\gamma + 2}$$

Với σ_{false} :

$$t^* = \frac{\int_0^1 [(\mu_{\sigma_{false}}(t))^\gamma t] dt}{\int_0^1 (\mu_{\sigma_{false}}(t))^\gamma dt} = \frac{\int_0^1 [(1-t)^\beta]^\gamma t dt}{\int_0^1 ((1-t)^\beta)^\gamma dt} = \frac{\int_0^1 [(1-t)^\beta]^\gamma t dt}{\int_0^1 ((1-t)^\beta)^\gamma dt} = \frac{\int_0^1 [(1-t)^{\beta\gamma}] t dt}{\int_0^1 (1-t)^{\beta\gamma} dt}$$

$$= \frac{\int_0^1 [(1-t)^{\beta\gamma} \cdot (1-(1-t))] dt}{\int_0^1 (1-t)^{\beta\gamma} dt} = 1 - \frac{\int_0^1 [(1-t)^{\beta\gamma+1} dt]}{\int_0^1 (1-t)^{\beta\gamma} dt} = 1 - \frac{\beta\gamma+1}{\beta\gamma+2} = \frac{1}{\beta\gamma+2}$$

3.2.5.3. Phương pháp khử mờ tam giác

Áp dụng công thức khử mờ tổng quát:

Với ***σtrue***, có các trường hợp $a \geq 0$ và $a \leq 0$.

Trường hợp $a \geq 0$:

$$t^* = \frac{\int_a^1 [(\frac{t-a}{1-a})^\gamma \cdot t] dt}{\int_a^1 [(\frac{t-a}{1-a})^\gamma] dt} = \frac{\gamma + a + 1}{\gamma + 2} \quad (3.14)$$

Trường hợp $a \leq 0$:

$$t^* = \frac{\int_a^1 [(\frac{t-a}{1-a})^\gamma \cdot t] dt}{\int_a^1 [(\frac{t-a}{1-a})^\gamma] dt} = \frac{(\gamma + a + 1)(1-a)^{\gamma+1} - a(-a)^{\gamma+1}}{(\gamma + 2)(1-a)^{\gamma+1} - (\gamma + 2)(-a)^{\gamma+1}} \quad (3.15)$$

Nếu $a = 0$ (tương ứng với *true*) thì: $t^* = \frac{\gamma + 1}{\gamma + 2}$

Với ***σfalse***, có các trường hợp $a \leq 1$ và $a \geq 1$.

Trường hợp $a \leq 1$:

$$t^* = \frac{\int_0^a [(1-\frac{t}{a})^\gamma \cdot t] dt}{\int_0^a (1-\frac{t}{a})^\gamma dt} = \frac{a}{\gamma + 2} \quad (3.16)$$

Trường hợp $a \geq 1$:

$$t^* = \frac{\int_0^1 [(1-\frac{t}{a})^\gamma \cdot t] dt}{\int_0^1 (1-\frac{t}{a})^\gamma dt} = \frac{(\gamma + a + 1)(a-1)^{\gamma+1} - a(a)^{\gamma+1}}{(\gamma + 2)(a-1)^{\gamma+1} - (\gamma + 2)(a)^{\gamma+1}} \quad (3.17)$$

3.2.6. Lập luận xấp xỉ mờ

L.A.Zadeh đã phát triển lý thuyết lập luận xấp xỉ (lập luận mờ) nhằm đưa ra một phương pháp biểu diễn và lập luận với các thông tin ngôn ngữ không chính xác.

Trong tiếp cận của Zadeh, lập luận xấp xỉ dựa trên cơ sở logic mờ giá trị ngôn ngữ. Các giá trị chân lý của mệnh đề là các giá trị ngôn ngữ của biến TRUTH.

Mỗi giá trị chân lý ngôn ngữ có ngữ nghĩa tương ứng với một tập mờ xác định trong đoạn $[0, 1]$ [28, 30].

Giả sử V là một biến có thể nhận các giá trị trên tập cơ sở X (giả sử X hữu hạn). Giá trị biến V có thể được biểu diễn bởi một tập mờ A của X trong mệnh đề dữ liệu có dạng: “ V is A ”. Ví dụ như: V is tall

Giả sử hai mệnh đề “ V_1 is A ” và “ V_2 is B ”, trong đó A, B là hai tập mờ tương ứng của X và Y .

Khi đó, mệnh đề hội: “ V_1 is A and V_2 is B ” cảm sinh một quan hệ mờ D trên tập tích Descartes $X \times Y$ và được biểu diễn hình thức như: “ (V_1, V_2) is D ”. (V_1, V_2) được gọi là một biến nối (*joint variable*).

Mệnh đề tuyển: “ V_1 is A or V_2 is B ” cảm sinh một quan hệ mờ E trên $X \times Y$.

Phép hội và phép tuyển của các mệnh đề được chuyển dịch trong lý thuyết lập luận xấp xỉ của Zadeh như sau:

$$D_{A \wedge B}(x, y) = \min[A(x), B(y)] \quad (3.18)$$

$$D_{A \vee B}(x, y) = \max[A(x), B(y)] \quad (3.19)$$

Mệnh đề kéo theo (*implication proposition*): “IF V is A THEN U is B ” cảm sinh quan hệ mờ trên $X \times Y$ được xác định theo Zadeh như sau:

$$F_{B/A(x,y)} = \min[1, 1 - A(x) + B(y)] \quad (3.20)$$

Luật suy diễn được sử dụng nhiều nhất trong lập luận xấp xỉ là suy diễn hợp thành mờ (*fuzzy copostinal iference*) và có ứng dụng quan trọng trong tình huống sau:

V is A'

IF V is A THEN U is B

U is B'

Trong trường hợp này luật suy diễn cho phép suy ra mệnh đề “ U is B ” bởi:

$$B' = A' \circ F_{B/A}$$

◦ – Ký hiệu hợp thành max – min, được Zadeh định nghĩa như sau:

$$B'(y) = \max \left[\min \left[A'(x), F_{B/A}(x, y) \right] \right]$$

Quan hệ mờ F cảm sinh bởi mệnh đề điều kiện: “ V is A THEN U is B ”, ký hiệu: $F_{A \rightarrow B}(x, y) = f \rightarrow [A(x), B(y)]$

Trong trường hợp tổng quát: $B' = A' \bullet F_{A \rightarrow B}$. Trong đó \bullet là ký hiệu hợp thành max-T, định nghĩa như sau: $B'(y) = \max \{T[A'(x), F_{A \rightarrow B}(x, y)], x \in X\}$

Việc suy diễn mờ có thể xem là sự khái quát hoá của luật Modus Ponens, từ đó có thể sử dụng hàm biến đổi giá trị chân lý để ước lượng tập mờ D .

Tập giá trị chân lý có một ý nghĩa quan trọng đối với logic mờ mà để xử lý, thông thường phải chuyển thành các tập mờ.

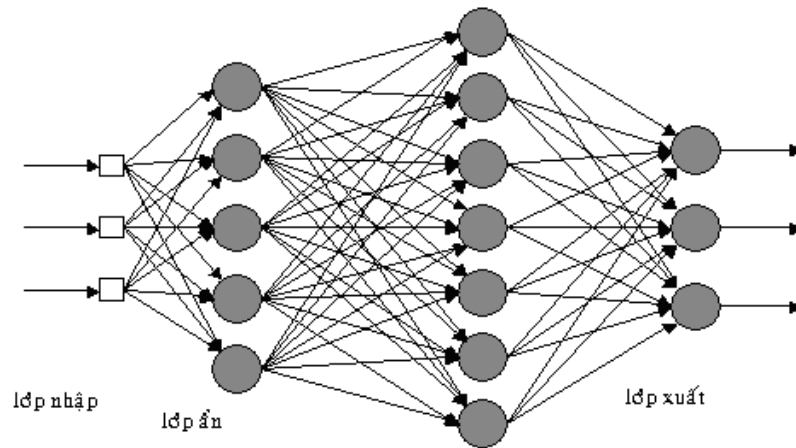
Từ cách biểu diễn giá trị chân lý ngôn ngữ thực hiện một ánh xạ định lượng ngữ nghĩa, và từ đó có thể áp dụng các phương pháp lập luận xấp xỉ.

3.3. Mạng nơron nhân tạo

3.3.1. Giới thiệu về mạng nơron nhân tạo

Xét về mặt cấu trúc, mạng nơron ANN (*Artificial Neural Network*) có cấu trúc phân lớp. Các cung được nối từ một nút ở lớp này đến các nút ở lớp kế tiếp. Hai nút trong cùng một lớp thì không kết nối với nhau. Mỗi nút trong một lớp nhận giá trị từ các nút ở lớp liền trước, tổng hợp lại theo trọng số của cung kết nối và chuyển giá trị kết xuất của nó cho các nút ở lớp liền sau.

Lớp đầu tiên nhận giá trị từ bên ngoài vào và được gọi là lớp nhập (*input*). Các nút trong lớp nhập được gọi là nút nhập. Lớp cuối cùng sẽ xuất ra kết quả của mạng và được gọi là lớp xuất (*output*). Các nút trong lớp xuất được gọi là nút xuất. Các lớp còn lại được gọi là lớp ẩn và các nút tương ứng được gọi là nút ẩn [27].



Hình 3.8. Minh hoạ mô hình mạng nơron 4 lớp.

Mô hình mạng nơron 4 lớp: 1 lớp nhập, 2 lớp ẩn và một lớp xuất. Quá trình phân lớp là quá trình ánh xạ một mẫu x từ không gian các đặc trưng vào không gian các lớp. Cũng như vậy, ANN thực chất là một hàm ánh xạ một vector đầu vào x thành một vector đầu ra z . Hàm này có đặc tính sau:

- + Là hàm phi tuyến (*nonlinear*).
- + Có tính ổn định (*stable*). Nếu một giá trị x_0 được ánh xạ thành một giá trị z_0 thì một giá trị x_1 “gần” với x_0 sẽ được ánh xạ thành một giá trị y_1 gần với y_0 . Mạng nơron cũng có thể ánh xạ một giá trị x_2 và x_0 thành một giá trị y_2 gần với y_0 .

Thực tế mạng nơron là một bộ máy nội suy và ngoại suy phi tuyến. Một mạng chỉ với một lớp ẩn là đã có thể xấp xỉ bất cứ một hàm phi tuyến nào thông qua một số mẫu trong tập mẫu.

Để đạt được điều này, cần luyện mạng bằng cách thay đổi các trọng số để ánh xạ từ các giá trị trong tập mẫu đến các giá trị đích mong muốn. Quá trình luyện mạng này cần có tập các vector mẫu đầu vào và đầu ra mong muốn tương ứng. Do đó, quá trình học này là quá trình học có giám sát.

3.3.2. Hoạt động của mạng nơron nhân tạo

Mạng nơron ANN chỉ có thể ở một trong hai trạng thái: Trạng thái ánh xạ và trạng thái học.

Ở trạng thái ánh xạ: Thông tin sẽ được lan truyền tiến từ các nút nhập đến các nút xuất và một mẫu x sẽ được ánh xạ thành một kết quả z .

Ở trạng thái học: Các trọng số của kết nối sẽ được điều chỉnh theo một thuật toán học để mạng có thể xấp xỉ được một hàm mong muốn nào đó. Thuật toán lan truyền ngược là một thuật toán hiệu quả cho quá trình học của ANN [31].

3.3.2.1. Trạng thái ánh xạ của mạng nơron nhân tạo

Ở trạng thái ánh xạ, mỗi vector đầu vào x sẽ được ánh xạ thành một vector kết quả z . Quá trình này được thực hiện như sau: Đầu tiên vector mẫu x sẽ được đưa vào lớp nhập. Mỗi nơron trong lớp nhập sẽ mang giá trị của một thành phần của x . Các nút nhập sẽ không tính toán gì cả mà gửi trực tiếp giá trị của nó đến các nơron ở lớp tiếp theo. Tại mỗi nơron của các lớp tiếp theo, một thao tác giống nhau sẽ được thực hiện. Đầu tiên nó sẽ tính tổng trọng hóa của tất cả các giá trị được gửi tới. Sau đó, một hàm truyền sẽ được áp dụng trên tổng trọng hóa này để cho giá trị xuất của nút này. Hàm truyền có tác dụng nén giá trị của tổng trọng hóa vào một miền giới hạn nào đó. Giá trị này được truyền cho các nơron ở lớp kế tiếp. Cứ thế thông tin được lan truyền cho đến lớp xuất của mạng [27].

Để đơn giản, thực hiện khảo sát mạng gồm 3 lớp: 1 lớp vào, 1 lớp ẩn và 1 lớp xuất. Thực tế cũng chỉ cần mạng 3 lớp là đủ để xấp xỉ các loại hàm.

Đối với nút ẩn, tổng trọng gửi tới nút j là:

$$u_j = a_{0j} + \sum_{i=1}^I a_{ij} x_i \quad (3.21)$$

Kết xuất của mạng có j nút ẩn là:

$$y_j = g(u_j) \quad j = 1, \dots, J \quad (3.22)$$

Trong đó i là số nút nhập x_i , a_{ij} là các trọng từ input i đến nút ẩn j và a_{0j} là trọng ngưỡng của nút ẩn j , $g(x)$ là hàm truyền.

Đối với mạng K nút xuất:

Tổng trọng gửi tới nút xuất k là:

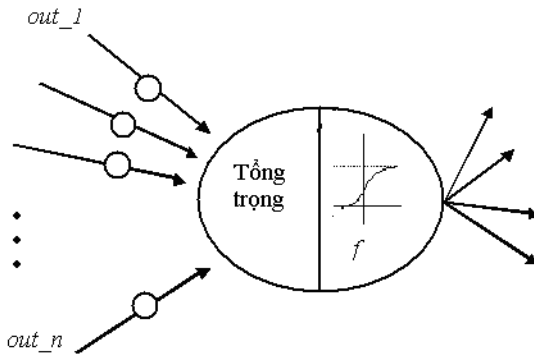
$$v_k = b_{0k} + \sum_{j=1}^J b_{jk} y_j \quad (3.23)$$

Kết xuất của mạng:

$$z_k = g(v_k) \quad k = 1, \dots, K \quad (3.24)$$

Trong đó j là số nút ẩn với các kết xuất y_j , b_{jk} là các trọng trên các cung liên kết từ nút ẩn j đến nút xuất thứ k , còn b_{0k} là trọng ngưỡng của nút xuất thứ k , $g(v_k)$ là hàm truyền theo k .

Ta có sơ đồ thể hiện các thao tác được thực hiện tại mỗi nơon.

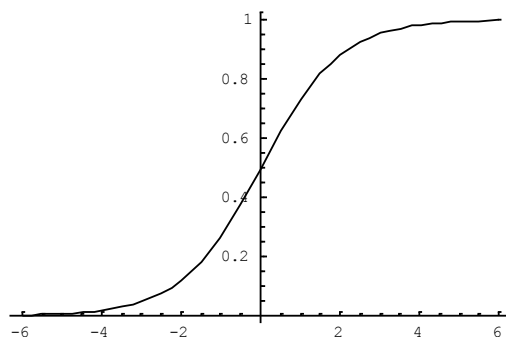


Hình 3.9. Mô tả hàm tính tổng trọng

Một số hàm truyền thường được sử dụng là:

+ Hàm sigmoid (hay hàm logistic) được xác định bởi:

$$f(x) = \frac{1}{1 + e^{-ax}} \quad (3.25)$$

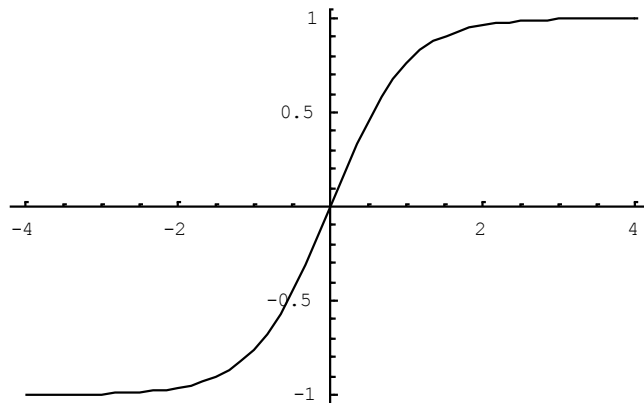


Hình 3.10. Mô tả hàm truyền sigmoid

Hình vẽ đồ thị hàm logistic ($a = 1$). Miền giá trị của hàm là $(0, 1)$.

+ Hàm tanh (tan- hyperbol)

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.26)$$



Hình 3.11. Hàm truyền tan- hyperbol

Đồ thị hàm tanh. Miền giá trị của hàm là $(-1, 1)$.

Hàm này có miền giá trị tương ứng là $(0, 1)$ và $(-1, 1)$. Việc sử dụng hàm logistic ($a = 1$) hay hàm tanh thực ra là tương đương với nhau vì chúng liên hệ tuyến tính với nhau. Việc sử dụng các hàm truyền khác nhau có liên hệ với các khoảng giá trị khác nhau của trọng số. Thực tế cho thấy, hàm tanh thường cho tốc độ hội tụ nhanh hơn trong quá trình học.

Các hàm truyền có thể được áp dụng vào các nút xuất hoặc không tùy vào mục đích ánh xạ của mạng. Nếu cần có một số giới hạn nhất định đối với đầu ra thì thực hiện áp dụng một hàm truyền thích hợp cho các nút xuất.

Khi mạng nơron được ứng dụng cho nhận dạng thì quá trình nhận dạng chính là quá trình ánh xạ của mạng nơron [29,31].

3.3.2.2. Trạng thái học của mạng nơron nhân tạo

Xét mạng ANN có một lớp ẩn với thuật toán lan truyền ngược.

Thuật toán lan truyền ngược là thuật toán hữu hiệu cho quá trình học của ANN. Thuật toán này sẽ cập nhật trọng số dựa trên một hàm lỗi E giữa kết xuất của mạng với giá trị đích.

Mục đích của việc học có giám sát bằng ANN là cực tiểu hoá hàm lỗi này. Kỹ thuật cơ bản để cực tiểu hoá hàm lỗi là phương pháp giảm gradient. Mặc dù phương pháp này có thể dẫn đến một cực tiểu cục bộ, nhưng nó được áp dụng rộng rãi vì tính đơn giản của nó. Thực tế cũng cho thấy trong hầu hết trường hợp phương pháp giảm gradient đều cho kết quả chấp nhận được [27, 31].

Quá trình học của mạng ANN theo thuật toán lan truyền ngược sẽ lặp đi lặp lại các thao tác sau:

- + Lan truyền tiến: Tính kết xuất y của mạng với một mẫu x .
- + Lan truyền ngược: Tính sai số giữa kết xuất y , giá trị đích t và lan truyền ngược sai số này lại để cập nhật trọng số cho mạng.

Quá trình học sẽ dừng khi mạng đã đạt được một độ lỗi nhỏ nhất định.

3.3.2.3. Phương pháp học giảm gradient

Phương pháp giảm gradient gồm các bước chính sau:

- + Chọn ngẫu nhiên một điểm x_0 trong không gian trọng số.
- + Tính độ dốc của hàm lỗi tại x_0 .
- + Di chuyển điểm x_0 theo hướng dốc nhất của hàm lỗi.

Quá trình tính độ dốc và di chuyển điểm x_0 được lặp đi lặp lại cho đến khi x_0 tiến đến giá trị làm cho hàm lỗi cực tiểu (có thể là hàm lỗi cực tiểu địa phương).

3.3.2.4. Cập nhật trọng số theo phương pháp giảm gradient

Phương pháp cập nhật trọng số theo hướng giảm gradient sẽ dựa trên đạo hàm riêng phần của hàm lỗi E đối với trọng số đang xét theo công thức sau:

$$w_{ij}(t+1) = w_{ij}(t) - \varepsilon \frac{\partial E}{\partial w_{ij}}(t) \quad (3.27)$$

Trong đó:

- + t là chỉ số của lần cập nhật trọng số hiện tại.
- + ε được gọi là hệ số học (*learning rate*).
- + w là một trọng số bất kì trong mạng: a_{ij} hoặc b_{ij}

Công thức trên có thể diễn dịch như sau: Cập nhật lại trọng số theo hướng ngược hướng của gradient với độ dài vector dịch chuyển phụ thuộc vào ε và độ lớn của vector đạo hàm.

Nếu trọng số được cập nhật theo hướng ngược với gradient một độ dịch chuyển vừa phải thì giá trị của hàm lỗi sẽ giảm đi so với trước khi cập nhật.

Giá trị của ε có ảnh hưởng lớn đến tốc độ hội tụ của thuật toán. Nếu ε lớn thì độ dịch chuyển lớn, kết xuất của mạng có thể dao động rất thất thường (vì có thể nhảy qua điểm cực tiểu) và thuật toán khó hội tụ. Nếu ε nhỏ thì phải cần rất nhiều bước lặp để đi đến được vị trí cực tiểu của hàm lỗi.

3.3.2.5. Quy tắc tính đạo hàm hàm lỗi

Sai số trung bình bình phương thường được sử dụng để đo lường sự trùng khớp giữa ánh xạ (ký hiệu NN) cần xây dựng với hàm đích cho trước (qua tập mẫu).

Cho tập mẫu

$$\Omega = \left\{ (X_k, Z_k) = (x_{k1}, \dots, x_{kM}; z_{k1}, \dots, z_{kN}); x_{ki}, z_{kj} \in R; i = 1..M; j = 1..N; k = 1..K \right\}$$

gọi $T_k = NN(X_k) = (t_{k1}, \dots, t_{kN})$ thì sai số trung bình bình phương sẽ là:

$$E = \frac{\frac{1}{2} \sum_{n=1}^N \sum_{k=1}^K (z_{kn} - t_{kn})^2}{N.K} \quad (3.28)$$

Đạo hàm hàm lỗi có thể được tính dựa vào quy tắc chuỗi như sau:

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial w_1} \frac{\partial w_1}{\partial w_2} \dots \frac{\partial w_n}{\partial w} \quad (3.29)$$

Xét từng trường hợp cụ thể:

Trọng số nút xuất: Chuỗi tính đạo hàm hàm lỗi theo trọng số giữa lớp ẩn và lớp xuất:

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial z} \frac{\partial z}{\partial v} \frac{\partial v}{\partial b} \quad (3.30)$$

Một nút xuất không ảnh hưởng gì đến sai số của các nút xuất khác trong lớp xuất. Để đơn giản trong công thức, bỏ qua chỉ số của các nút xuất, nếu nút xuất đang xét có giá trị thực là z và giá trị đúng của nút đó là t , thì sai số bình phương là:

$$E = \frac{1}{2}(z - t)^2 \quad (3.31)$$

Từ đó, ta có:

$$\frac{\partial E}{\partial z} = z - t \quad (3.32)$$

Mặt khác độ dốc hàm logistic được tính như sau:

$$g(v)' = \frac{\partial z}{\partial v} = z(1 - z) \quad (3.33)$$

Tính số hạng thứ 3, $\partial v / \partial b$ trong công thức $\partial E / \partial b$

$$v = b_0 + \sum_{j=1}^J b_j y_j \quad (3.34)$$

Xét với b_0 thì $\partial v / \partial b_0 = 1$, với trọng số b_j thì $\partial v / \partial b_j = y_j$
Đặt:

$$p \equiv \frac{\partial E}{\partial z} \frac{\partial z}{\partial v} \quad (3.35)$$

Ta có:

$$p = (z - t)z(1 - z) \quad (3.36)$$

Như vậy:

$$\frac{\partial E}{\partial b_j} = \begin{cases} p & j = 0 \\ py_j & j \leq J \end{cases} \quad (3.37)$$

Xét trọng số nút ẩn: Đạo hàm của hàm lỗi đối với các trọng số của một nút ẩn:

$$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial u} \frac{\partial u}{\partial a} \quad (3.38)$$

Các nút ẩn tự chúng không tạo lỗi nhưng chúng góp phần tác động vào lỗi của các nút xuất. Những tác động này cũng xác định theo luật chuỗi:

$$\frac{\partial E}{\partial y} = \sum_{k=1}^K \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial v_k} \frac{\partial v_k}{\partial y} \quad (3.39)$$

(trong công thức này chỉ số j đã được bỏ đi vì chỉ có một nút ẩn tham gia)

Công thức trên cho thấy ảnh hưởng của nút ẩn vào hàm E là một tổng theo tất cả các nút xuất ($k = 1 \dots K$) của tích ba số hạng.

Hai số hạng đầu trong tích đó có ý nghĩa tương tự như phần trên. Ở trên đặt tích này là p . Lượng p này lan truyền ngược từ nút xuất đến nút ẩn. Nếu cho p một chỉ số cho biết nó thuộc về nút xuất nào, có thể viết phương trình trên lại như sau:

$$\frac{\partial E}{\partial y} = \sum_{k=1}^K p_k \frac{\partial v_k}{\partial y} \quad (3.40)$$

Xét số hạng $\partial v_k / \partial y$, ta có:

$$\frac{\partial v_k}{\partial y} = b_k \quad (3.41)$$

Như vậy:

$$\frac{\partial E}{\partial y} = \sum_{k=1}^K p_k b_k \quad (3.42)$$

Ta tiếp tục xét $\partial y / \partial u$ trong chuỗi tính $\partial E / \partial a$

$$\frac{\partial y}{\partial u} = y(1 - y) \quad (3.43)$$

Đặt:

$$q \equiv \frac{\partial E}{\partial y} \frac{\partial y}{\partial u} \quad (3.44)$$

Ta có:

$$q = \left(\sum_{k=1}^K p_k b_k \right) y(1 - y) \quad (3.45)$$

Và cuối cùng $\partial u / \partial a$, trong chuỗi tính $\partial E / \partial a$ chính là thay đổi của u đối với thay đổi của một trọng số. Nó phụ thuộc vào loại trọng số.

Đối với trọng ngưỡng a_0 :

$$\frac{\partial u}{\partial a_0} = 1 \quad (3.46)$$

Với các trọng trên cung nối trực tiếp a_i , $i > 0$:

$$\frac{\partial u}{\partial a_i} = x_i \quad (3.47)$$

Tóm lại:

$$\frac{\partial E}{\partial a_i} = \begin{cases} q & i = 0 \\ qx_i & i > 0 \end{cases} \quad (3.48)$$

3.3.3. Một số kỹ thuật huấn luyện mạng

3.3.3.1. Học theo lô

Trọng số của mạng được cập nhật khi tất cả các mẫu học đều đã được đưa vào mạng (học theo bước học). Đây là một kỹ thuật học theo lô.

Trong cách học thông thường của thuật toán lan truyền ngược, mỗi khi ta đưa vào mạng một mẫu học đạo hàm hàm lỗi được tính và trọng số được cập nhật ngay, sau đó mẫu học tiếp theo được đưa vào mạng và quá trình trên được lặp lại cho từng mẫu học. Mỗi bước trong quá trình trên gọi là một bước lặp. Khi tất cả các mẫu trong tập học đã được đưa vào mạng, ta hoàn tất bước học. Cách cập nhật này có thể dẫn đến trường hợp ở một bước lặp p trọng số được cập nhật sao cho giảm $E(p)$, nhưng ở một bước lặp q khác, trọng số lại được cập nhật sao cho tăng $E(q)$. Kết quả là một bước học sẽ chứa đựng sự tự mâu thuẫn trong nó, cụ thể sẽ có một bước lặp làm ngược lại với những kết quả đã đạt được ở những bước trước một cách cục bộ tại một trọng số. Điều này làm cho quá trình học nhiễu loạn và kém hiệu quả.

Cách học theo bước giúp ta tránh được hiện tượng này. Trong phương pháp này trọng số sẽ được cập nhật sau mỗi bước học. Nghĩa là tất cả các mẫu học sẽ được đưa vào mạng và các đạo hàm riêng phần ứng với mỗi trọng số và mỗi mẫu học sẽ được tính toán. Sau đó đạo hàm riêng phần ứng với mỗi trọng số sẽ được tính bằng tổng tất cả các đạo hàm riêng phần ứng với trọng số đó trên tất cả các mẫu học.

Phương pháp học theo bước học đã được ứng dụng rộng rãi trong các hệ thống xử lý lớn nhờ khả năng áp dụng việc xử lý song song cho việc tính toán các đạo hàm riêng phần [27, 29].

3.3.3.2. Ngăn chặn quá khớp

Quá khớp là trường hợp mạng thuộc hết dữ liệu học (kể cả nhiễu), lúc đó nó sẽ trả lời chính xác những gì nó được học còn những gì nó không được học thì không quan tâm. Nghĩa là mạng không có khả năng tổng quát hóa, mà tổng quát lại là điều ta cần khi sử dụng mạng.

Có nhiều cách khắc phục tình trạng này song cách đơn giản nhất là chia mẫu thành tập luyện và tập kiểm tra. Luyện mạng với tập mẫu luyện, nhưng định kỳ ngừng luyện để đánh giá sai số bằng tập mẫu kiểm tra. Mỗi lần ngừng lại để đánh giá sai số trên mẫu kiểm tra, cần lưu lại các trọng số. Khi sai số trên mẫu kiểm tra đi lên, thì quá khớp đã bắt đầu. Do vậy, ngừng luyện, trở về trọng sinh lỗi thấp nhất trên mẫu kiểm tra [27, 31].

3.3.4. Dự báo dữ liệu chuỗi thời gian

Dựa trên mạng nơron đã được huấn luyện từ tập dữ liệu quá khứ, ứng với dữ liệu tương lai (dữ liệu mới) sẽ thực hiện dự báo giá trị đầu ra tương ứng. Do đó, mạng nơron có thể thực hiện trong bài toán dự báo giá cổ phiếu trong tương lai nhằm giúp cho các nhà kinh doanh có được định hướng trong chiến lược của mình. Mỗi dữ liệu trong quá khứ được đo đạc theo thực tế và có cấu trúc là một ma trận các giá trị thực với mỗi dòng mô tả một mã cổ phiếu ứng với bốn trạng thái <Open>, <High>, <Low>, <Close>. Do đó, giá cổ phiếu phụ thuộc vào từng ngày diễn ra các phiên giao dịch.

Giả sử khung thời gian trong trường hợp này là $\Delta t = 1$ ngày, dữ liệu này có thể được mô tả là một dữ liệu chuỗi thời gian $X_1 X_2 \dots X_t$. Vì vậy, ta có thể sử dụng mạng nơron để dự báo dữ liệu tương lai gồm X_{t+1}, X_{t+2}, \dots

3.4. Tổng kết chương

Trong chương này đã khảo sát hai đối tượng nghiên cứu chính của luận văn là hệ suy luận mờ và mạng nơron nhân tạo. Bên cạnh việc giới thiệu về các tham số trong mô hình suy luận mờ, nội dung của chương cũng tập trung vào việc nghiên cứu các phương pháp suy luận mờ. Luận văn sẽ áp dụng phương pháp suy diễn mờ để tìm ra các tập luật mờ nhằm thực hiện quá trình suy luận các kết quả dữ liệu theo chuỗi thời gian. Nội dung của chương tiếp cận mô hình mạng nơron nhân tạo để dự báo dữ liệu chuỗi thời gian. Ngoài ra, luận văn sẽ sử dụng mạng neuro-fuzzy ANFIS, tức là kết hợp giữa hệ suy luận mờ và mạng nơron nhân tạo để giải quyết bài toán dự báo giá cổ phiếu.

CHƯƠNG 4

HỆ SUY LUẬN ANFIS

4.1. Giới thiệu

Trong chương này tiếp cận mô hình hệ suy luận ANFIS để từ đó đề xuất mô hình dự báo neuro-fuzzy ANFIS đa trị. Để thực hiện được vấn đề này, trong chương này lần lượt nghiên cứu về hệ suy luận ANFIS truyền thống để từ đó thiết kế mô hình suy luận đa trị nhằm dự báo giá cổ phiếu trong tương lai bao gồm giá mở <Open>, giá cao nhất <High>, giá thấp nhất <Low>, giá đóng <Close> của các mã cổ phiếu. Dựa trên mô hình đã đề xuất, trong chương này trình bày thuật toán huấn luyện mạng neuro-fuzzy ANFIS nhằm tìm ra ma trận hệ số suy luận mờ. Từ đó nội dung của chương cũng trình bày thuật toán kiểm thử để kiểm chứng dữ liệu nhằm minh chứng quá trình dự báo cho người dùng.

4.2. Xây dựng mô hình suy luận mờ ANFIS nhằm dự báo giá cổ phiếu

4.4.1. Mô tả dữ liệu đầu vào

Luận văn thực hiện dự báo giá cổ phiếu cho từng loại cổ phiếu khác nhau. Mỗi loại cổ phiếu sẽ có mã cổ phiếu riêng biệt. Ứng với mỗi mã cổ phiếu sẽ có các trạng thái giao dịch tương ứng.

Không làm mất tính toàn vẹn của dữ liệu, giả sử xét khung thời gian $\omega = 1$, tức là khoảng cách giữa phiên giao dịch lần thứ nhất đã thông báo giá đóng với phiên giao dịch lần thứ hai khi đã thông báo giá đóng là 1 ngày.

Xét dữ liệu giao dịch tại ngày thứ i với m mã cổ phiếu, mỗi mã cổ phiếu có n trạng thái tương ứng. Vì vậy, bộ dữ liệu đầu vào tại ngày thứ i là một ma trận như sau:

$$X_i = \begin{bmatrix} x_{11}^{(i)} & x_{12}^{(i)} & \dots & x_{1n}^{(i)} \\ x_{21}^{(i)} & x_{22}^{(i)} & \dots & x_{2n}^{(i)} \\ \dots & \dots & \dots & \dots \\ x_{m1}^{(i)} & x_{m2}^{(i)} & \dots & x_{mn}^{(i)} \end{bmatrix} \quad (4.1)$$

Luận văn sẽ trình bày quá trình dự báo với m mã cổ phiếu trong tương lai từ bộ dữ liệu đầu vào gồm N bộ dữ liệu, mỗi bộ dữ liệu là 1 ngày mô tả cho giao dịch của m mã cổ phiếu. Khi đó, bộ dữ liệu đầu vào là:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \dots & x_{1n}^{(1)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \dots & x_{2n}^{(1)} \\ \dots & \dots & \dots & \dots \\ x_{m1}^{(1)} & x_{m2}^{(1)} & \dots & x_{mn}^{(1)} \\ x_{11}^{(2)} & x_{12}^{(2)} & \dots & x_{1n}^{(2)} \\ x_{21}^{(2)} & x_{22}^{(2)} & \dots & x_{2n}^{(2)} \\ \dots & \dots & \dots & \dots \\ x_{m1}^{(2)} & x_{m2}^{(2)} & \dots & x_{mn}^{(2)} \\ \dots & \dots & \dots & \dots \\ x_{11}^{(N)} & x_{12}^{(N)} & \dots & x_{1n}^{(N)} \\ x_{21}^{(N)} & x_{22}^{(N)} & \dots & x_{2n}^{(N)} \\ \dots & \dots & \dots & \dots \\ x_{m1}^{(N)} & x_{m2}^{(N)} & \dots & x_{mn}^{(N)} \end{bmatrix} \quad (4.2)$$

4.4.2. Xây dựng mô hình ANFIS

Dữ liệu đầu vào là một ma trận mô tả giá cổ phiếu. Mỗi dòng của ma trận là một vector gồm $m \times n$ giá trị rõ. Thực hiện mờ hóa các giá trị rõ. Dựa vào các giá trị mờ, tính tích trọng số mờ. Sau đó, tiến hành chuẩn hóa các trọng số mờ để tạo giá trị đầu vào cho bài toán tích hợp mờ. Tổng hợp các giá trị mờ chính là giá trị dự đoán trong tương lai.

Luận văn đề xuất mô hình ANFIS dự báo giá cổ phiếu (**Hình 4.1**) cho N bộ dữ liệu X_1, X_2, \dots, X_N . Mỗi bộ dữ liệu X_i là một vector mô tả giá cổ phiếu tại ngày thứ i của m mã cổ phiếu với n trạng thái tương ứng $x_{11}^{(i)}, x_{12}^{(i)}, \dots, x_{1n}^{(i)}, x_{m1}^{(i)}, x_{m2}^{(i)}, \dots, x_{mn}^{(i)}$. Vì vậy, tại tầng 1 của mô hình gồm N bộ dữ liệu mô tả giá cổ phiếu theo chuỗi thời gian liên tục. Đầu ra của tầng 1 là các giá trị mờ ứng với giá trị rõ trên từng bộ dữ liệu giá cổ phiếu. Các giá trị mờ ứng với chuỗi dữ liệu giá cổ phiếu trong N ngày liên tục được tích hợp và trở thành đầu vào của tầng 2.

Kết quả đầu ra của tầng 2 là một vector như sau:

$$W = (w_1^{(1)}, w_2^{(1)}, \dots, w_n^{(1)}, w_1^{(2)}, w_2^{(2)}, \dots, w_n^{(2)}, w_1^{(N)}, w_2^{(N)}, \dots, w_n^{(N)}) \quad (4.3)$$

Trong đó, $w_j^{(N)} (j=1..n)$ là giá trị tích hợp từ N thành phần mờ ứng với n trạng thái của chuỗi dữ liệu cổ phiếu.

Tại tầng 3, thực hiện chuẩn hóa vector W để tạo thành:

$$\bar{W} = (\overline{w_1^{(1)}}, \overline{w_2^{(1)}}, \dots, \overline{w_n^{(1)}}, \overline{w_1^{(2)}}, \overline{w_2^{(2)}}, \dots, \overline{w_n^{(2)}}, \overline{w_1^{(N)}}, \overline{w_2^{(N)}}, \dots, \overline{w_n^{(N)}}) \quad (4.4)$$

Trong đó, $\overline{w_j^{(N)}} \in [0,1]$.

Dựa trên phương pháp suy luận mờ Takagi-Sugeno, tại tầng 4 thực hiện quá trình suy luận để tìm ra các giá trị của ma trận dự báo

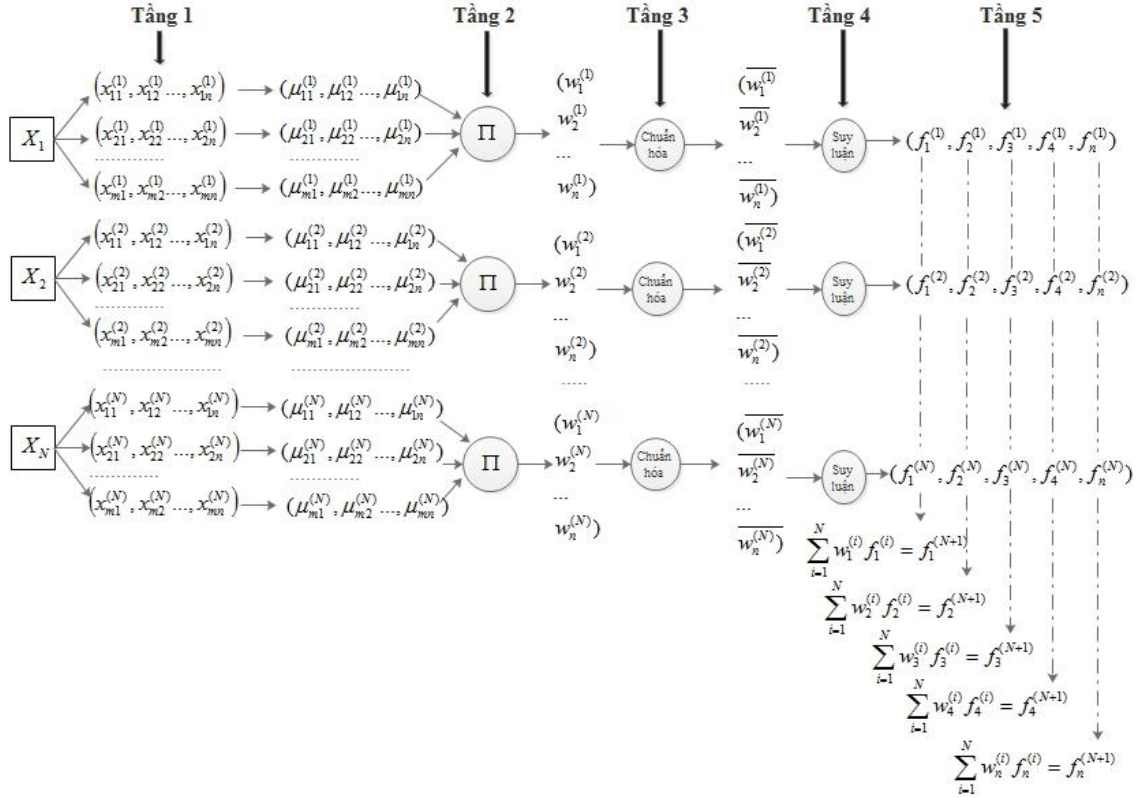
$$f_n^{(N)} = r_{N0} + r_{N1} \times Y_n^{(1)} + r_{N2} \times Y_n^{(2)} + r_{N3} \times Y_n^{(3)} + \dots + r_{Nn} \times Y_n^{(N)} \quad (4.5)$$

Trong đó, $Y_n^{(N)} = (x_{1n}^{(N)}, x_{2n}^{(N)}, \dots, x_{nn}^{(N)})$ tương ứng để làm đầu vào cho tầng 5 nhằm đưa ra giá trị dự báo.

Thực hiện huấn luyện ma trận hệ số suy luận $R_{m \times (m+1)} = (r_{ij})$, $i=1..m$ và $j=0..m$ cho đến khi ma trận hệ số suy luận hội tụ ứng với sai số ε cho trước thì dừng lại.

Mô hình được đề xuất có thể thực hiện dự báo đồng thời một ma trận kết quả tại cùng một thời điểm. Đây chính là đóng góp của luận văn, giúp cải thiện đáng kể về mặt tốc độ so với hệ ANFIS truyền thống.

Mô hình này áp dụng được dữ liệu dạng ma trận các số mờ, áp dụng được với từng khoảng dữ liệu liên tục (tức là dự báo cho các giá trị đơn lẻ, vô hướng). Vì vậy, mô hình này hoàn toàn thỏa mãn hệ ANFIS truyền thống.



Hình 4.1. Mô hình ANFIS dự báo giá cổ phiếu

4.4.3. Thuật toán huấn luyện hệ suy luận mờ ANFIS

Nhằm xây dựng hệ dự báo giá cổ phiếu dựa trên mạng ANFIS như đã đề xuất, bước đầu tiên cần phải thực hiện huấn luyện mạng ANFIS này. Kết quả huấn luyện sẽ có được một ma trận hệ số để từ đó giúp cho quá trình nội suy bằng phương pháp Takagi-Sugeno để có được kết quả dự báo. Trong phần này trình bày thuật toán huấn luyện mạng neuro-fuzzy ANFIS như đã đề xuất với đầu vào là tập dữ liệu huấn luyện được lấy từ dữ liệu thực tế trên Cophieu68. Để huấn luyện theo các giá trị thực tế, trong thuật toán huấn luyện này cần có một tập các vector giá trị thực để từ đó tính sai số trong mỗi lần lan truyền xuôi trên mạng neuro-fuzzy ANFIS. Dựa trên kết quả lan truyền xuôi, thuật toán huấn luyện thực hiện tính sai số, nếu như sai số chưa thỏa mãn điều kiện hội tụ thì thuật toán huấn luyện thực hiện quá trình lan truyền ngược để tinh chỉnh các hệ số suy luận sao cho phù hợp với giá trị thực tế, tức là làm giảm đi sai số dự báo.

Thuật toán huấn luyện mạng neuro-fuzzy ANFIS được trình bày như sau:

Đầu vào: Tập ma trận dữ liệu huấn luyện $Training = \{M_1, \dots, M_T\}$

Tập vector giá trị thực $Real = \{r_1, \dots, r_T\}$

Đầu ra: Tập ma trận hệ số suy luận $R = \{R_1, R_2, \dots, R_T\}$

Phương pháp

Bước 1: Khởi tạo

(1) Tạo tập ma trận mờ $M-fuzzy = \{Mf_1, \dots, Mf_T\}$ từ tập huấn luyện $Training$;

(2) Tạo tập vector trọng số $W = \{W_1, \dots, W_T\}$ từ tập ma trận mờ $M-fuzzy$;

(3) Chuẩn hóa tập vector trọng số W ;

(4) Khởi tạo ngẫu nhiên ma trận hệ số suy luận mờ R ;

(5) Tính sai số $Error$ ban đầu;

Bước 2: Huấn luyện (tức là cập nhật ma trận hệ số suy luận R)

(6) **Foreach** (với mỗi bộ dữ liệu huấn luyện Mf_i) **do**

(7) **While** ($Error > \varepsilon$ AND số lần lặp < 1000) **do**

(8) Tạo mới ma trận hệ số suy luận R_i dựa trên hàm lỗi;

(9) Tính ma trận dự báo F_i ;

(10) Chuẩn hóa ma trận dự báo;

(11) Tính kết quả dự báo;

(12) Tính sai số $Error$;

(13) **EndWhile**

(14) Lưu trữ ma trận hệ số R_i của bộ dữ liệu huấn luyện Mf_i ;

(15) **EndForeach**

Kết thúc

Trong thuật toán trên thực hiện gồm 2 bước, bước thứ nhất thực hiện quá trình khởi tạo các giá trị cần thiết cho quá trình huấn luyện, bước thứ hai thực hiện huấn luyện (tức là cập nhật ma trận hệ số suy luận R).

Tại dòng lệnh (1) của thuật toán huấn luyện thực hiện việc tính tập các ma trận mờ từ tập dữ liệu huấn luyện. Giả sử có T bộ dữ liệu huấn luyện, mỗi bộ dữ liệu huấn luyện có k trạng thái (trong thực nghiệm $k = 4$ ứng với 4 trạng thái giá cổ phiếu gồm giá mở <Open>, giá cao nhất <High>, giá thấp nhất <Low>, giá đóng <Close>). Ứng với mỗi bộ dữ liệu huấn luyện M_i , tại dòng lệnh (1) tính ma trận mờ tương ứng. Giả sử ma trận huấn luyện M_i có m dòng n cột và k trạng thái thì quá trình mờ hóa bao gồm $m \times n \times k$ phép toán mờ hóa. Do đó, độ phức tạp để mờ hóa một bộ dữ liệu M_i là $O(m \times n \times k \times C)$, trong đó C là hằng số mô tả số phép toán mờ hóa bằng hàm Gaussian. Vì vậy, độ phức tạp của dòng lệnh (1) là $O(T \times m \times n \times k \times C)$. Trong thực nghiệm thì $T = 50$, $m = 120$, $n = 30$, do đó độ phức tạp là $50 \times 120 \times 30 \times C$. Do đó, tại dòng lệnh thứ nhất độ phức tạp là tuyến tính $180,000 \times C$.

Tại dòng lệnh (2) thực hiện tính trọng số mờ ứng với mỗi ma trận mờ Mf_i để tạo thành các vector trọng số W_i . Trong trường hợp dự báo giá cổ phiếu của luận văn thì có 4 trạng thái, do đó mỗi vector trọng số W_i có 4 vector thành phần tương ứng với 4 trạng thái của giá cổ phiếu. Mỗi giá trị của một vector thành phần được tính bằng cách tổng hợp từng cột của ma trận hệ số mờ Mf_i . Do đó, cần phải thực hiện m phép toán. Tương tự như trên, độ phức tạp của dòng lệnh (2) trong trường hợp này là $O(T \times k \times m \times n)$. Vì các giá trị T, k, m, n là giới hạn, nên trong trường hợp này độ phức tạp của dòng lệnh (2) cũng là tuyến tính.

Tại dòng lệnh (3) thực hiện chuẩn hóa tập vector trọng số do đó độ phức tạp của dòng lệnh này sẽ là $O(T \times k \times n)$, với n là số cột của ma trận mờ. Như vậy, tại bước 1 độ phức tạp là $O(T \times k \times n)(m \times C + m + 1)$.

Tại bước 2 của thuật toán huấn luyện thực hiện quá trình huấn luyện trên mỗi bộ dữ liệu Mf_i . Tại dòng lệnh (7) thực hiện kiểm tra điều kiện hội tụ của quá trình huấn luyện của từng bộ dữ liệu. Mỗi lần huấn luyện cần thực hiện tối đa là V vòng lặp (trong thực nghiệm $V = 1,000$). Tại mỗi lần lặp, cần tính lại ma trận hệ số suy luận R_i tương ứng để từ đó tính ma trận dự báo F_i tại dòng lệnh (9). Độ phức tạp tại

dòng lệnh (8) và (9) là $O(k \times m \times n)$. Tại dòng lệnh (10) thực hiện chuẩn hóa ma trận dự báo F_i , tại dòng lệnh này độ phức tạp vẫn là $O(k \times m \times n)$. Dựa trên ma trận dự báo đã chuẩn hóa thực hiện tính kết quả dự báo bằng cách tổng hợp theo phương pháp suy luận mờ Takagi-Sugeno tại dòng lệnh (11), trong trường hợp này độ phức tạp là $O(m \times k)$. Tại dòng lệnh (12) thực hiện tính giá trị sai số bằng tổng trung bình độ lệch do đó độ phức tạp là $O(m \times k)$.

Trên cơ sở phân tích như trên, tại bước 2 của thuật toán huấn luyện có độ phức tạp là $O(V \times m \times k(3 \times n + 2))$.

Do đó, độ phức tạp của thuật toán huấn luyện là:

$$O(T \times k \times n)(m \times C + m + 1) + O(V \times m \times k(3 \times n + 2))$$

4.4.4. Thuật toán kiểm thử hệ suy luận mờ ANFIS

Sau khi thực hiện huấn luyện mạng neuro-fuzzy ANFIS, ta có được tập các ma trận hệ số suy luận. Trên cơ sở này, luận văn xây dựng thuật toán kiểm thử như sau:

Đầu vào: Ma trận dữ liệu kiểm thử M

Tập ma trận hệ số suy luận $R = \{R_1, R_2, \dots, R_n\}$

Đầu ra: Vector dự báo $F = \{f_1, f_2, \dots, f_m\}$, Vector sai số $Error = \{e_1, e_2, \dots, e_m\}$

Phương pháp

Bước 1: Khởi tạo

(1) Tạo ma trận hệ số mờ T từ ma trận M ;

(2) Tính vector trọng số W từ ma trận T ;

(3) Chuẩn hóa vector trọng số W ;

Bước 2: Dự báo

(4) Tính ma trận suy luận F ;

(5) Chuẩn hóa ma trận suy luận F ;

(6) Thực hiện tính kết quả dự báo;

(7) Tính sai số dự báo;

Kết thúc

Trong bước 1 thực hiện tạo ma trận hệ số mờ từ ma trận dữ liệu kiểm thử M tại dòng lệnh (1). Dựa trên ma trận hệ số mờ thực hiện tính vector trọng số W , trong vector trọng số này gồm có 4 vector thành phần mô tả trọng số của giá cổ phiếu ứng với 4 trạng thái. Từ đó, thực hiện chuẩn hóa vector trọng số.

Tại bước 2 thực hiện quá trình dự báo ứng với bộ dữ liệu đầu vào M . Tại dòng lệnh (4) thực hiện tính ma trận suy luận F và chuẩn hóa ma trận suy luận. Dựa trên phương pháp tổng hợp mờ Takagi-Sugeno và mô hình dự báo tại tầng 5, thực hiện tính kết quả dự báo đồng thời tính sai số dự báo tương ứng. Phương pháp dự báo dựa trên cơ sở thuật toán huấn luyện nhưng có dữ liệu đầu vào độc lập với dữ liệu huấn luyện, tức là chỉ thực hiện quá trình lan truyền xuôi, không thực hiện quá trình lan truyền ngược. Do đó, độ phức tạp của thuật toán dự báo trong trường hợp này là không đáng kể so với thuật toán huấn luyện, tức là:

$$O(k \times n)(m \times C + m + 1) + O(m \times k(3 \times n + 2))$$

4.3. Ví dụ thực nghiệm

Phần thực nghiệm dựa trên mô hình ANFIS được đề xuất tại **Hình 4.1** ứng với bộ dữ liệu được lấy từ cơ sở dữ liệu Cophieu68 từ ngày 10/04/2016 đến ngày 15/04/2016. Bộ dữ liệu thực nghiệm có 20 mã cổ phiếu. Mỗi mã cổ phiếu có bốn trạng thái: giá mở <Open>, giá cao nhất <High>, giá thấp nhất <Low>, giá đóng <Close> cùng với giá cổ phiếu tương ứng trong từng ngày.

Bảng 4.1. Ví dụ dữ liệu mẫu huấn luyện giá cổ phiếu

Mã CK	Ngày 1 D1				Ngày 2 D2				Ngày 3 D3				Ngày 4 D4				Ngày 5 D5			
	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
AAA	16.7	16.9	16.4	16.4	16.4	16.6	16.4	16.6	16.4	16.6	16.3	16.5	16.6	17.1	16.5	16.9	17	17	16.9	16.9
AAM	10.9	10.9	10.9	10.9	10.3	10.3	10.3	10.3	10.4	10.8	10.3	10.3	10.3	10.7	10.3	10.7	10.6	10.7	10.3	10.3
ACB	18.3	18.5	18.2	18.5	18.4	18.4	18.3	18.4	18.4	18.4	18.3	18.3	18.3	18.3	18	18.1	18	18.2	17.8	17.8
AGF	11.1	11.8	11.1	11.8	11.7	11.7	11.4	11.5	11.5	11.7	11.4	11.4	11.5	11.8	11.4	11.7	11.5	11.6	11.4	11.5
AGR	2.9	3	2.9	2.9	2.8	2.9	2.7	2.9	2.9	2.9	2.7	2.7	2.8	2.8	2.7	2.8	2.8	2.9	2.8	2.9
AMD	11.9	12.1	11.8	11.8	11.8	11.9	11.7	11.9	11.8	12	11.7	11.8	11.9	11.9	11.6	11.9	11.8	12.4	11.8	12.4
ASA	3	3.1	3	3.1	3.1	3.1	2.9	3	3	3.1	2.9	2.9	3	3.1	3	3.1	3.1	3.1	3	3.1
ASM	17.5	17.5	17.2	17.3	17.1	17.8	17.1	17.7	17.8	17.9	17.7	17.8	17.6	17.9	17.6	17.6	17.9	18.1	17.8	18
ASP	3.9	4	3.8	3.9	4.1	4.1	4.1	4.1	4.3	4.3	4.3	4.3	4.6	4.6	4.6	4.6	4.9	4.9	4.9	4.9
ATA	5.1	5.3	5.1	5.3	5.4	5.6	5.4	5.6	5.8	5.9	5.3	5.3	5.6	5.6	5.1	5.6	5.9	5.9	5.9	5.9
AVF	0.9	1	0.8	1	0.9	0.9	0.8	0.9	0.9	0.9	0.8	0.9	0.8	0.9	0.8	0.8	0.8	0.9	0.8	0.9
BAM	2.1	2.1	1.9	2	2.1	2.2	2	2.2	2.1	2.2	2	2	1.8	2	1.8	1.9	1.9	2	1.8	2

BCC	15.5	15.7	15.4	15.5	15.4	15.7	15.3	15.5	15.5	15.5	15.1	15.1	15.1	15.2	15	15.2	15	15.1
BCE	6.6	6.6	6.4	6.6	6.6	6.7	6.4	6.4	6.4	6.7	6.4	6.6	6.6	6.8	6.5	6.6	6.6	6.4
BCI	23.4	23.5	23.2	23.3	23.3	23.7	23.3	23.4	23.4	23.5	23.3	23.5	23.4	23.5	23.3	23.4	23.3	23.9
BGM	3.2	3.3	3.1	3.1	3.1	3.3	3.1	3.2	3.3	3.3	3	3	3	3.1	2.9	3	2.9	2.9
BHS	18.8	18.8	18.5	18.7	18.6	18.8	18.5	18.8	18.7	18.9	18.5	18.8	19.1	19.2	18.8	19.1	19.1	19.1
BIC	21.2	21.6	20.9	21.4	21.5	21.5	21	21.2	21.4	21.4	20.6	21.2	20.6	21.2	20.5	20.6	20.6	20.7
BID	16.9	16.9	16.8	16.8	16.8	16.9	16.5	16.7	16.5	16.6	16.1	16.2	15.8	16.1	15.5	15.8	15.8	15.9
BII	13	13.1	12.8	13	12.9	13.4	12.9	13.4	13.4	13.4	13.2	13.2	13.1	13.2	13	13	13	13.2

Theo **Bảng 4.1** mô tả dữ liệu thực nghiệm đối với các mã cổ phiếu như AAA, AAM,... trên dữ liệu thực tế từ Cophieu68. Mã cổ phiếu AAA được giao dịch trong 5 ngày và được đánh dấu từ X_1 đến X_5 . Mỗi ngày giao dịch sẽ bao gồm bốn trạng thái: O (<Open>), H (<High>), L (<Low>), C (<Close>), mỗi trạng thái sẽ có giá cổ phiếu tương ứng và đơn vị tính của giá cổ phiếu là 1,000 VNĐ.

Thực hiện mờ hóa bộ dữ liệu thực nghiệm bằng phương pháp Gaussian của tập mờ hình chuông. Khi đó, giá trị kỳ vọng c_i và độ lệch chuẩn a_i như sau:

Bảng 4.2. Giá trị tham số mờ hàm Gaussian

	D1				D2				D3				D4				D5			
	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
c_i	11.5	12.0	11.5	11.8	11.8	11.8	11.6	11.7	11.7	11.9	11.6	11.6	11.7	11.9	11.5	11.8	11.7	12.0	11.6	12.0
a_i	7.2	7.2	7.1	7.2	7.1	7.2	7.1	7.2	7.2	7.2	7.1	7.2	7.1	7.2	7.1	7.1	7.1	7.2	7.0	7.1

Theo **Bảng 4.2**, các giá trị kỳ vọng và độ lệch chuẩn ứng với từng trạng thái trong mỗi ngày. Dòng thứ nhất là giá trị kỳ vọng tương ứng với mỗi trạng thái của các tất cả các mã cổ phiếu. Dòng thứ hai là mô tả giá trị độ lệch chuẩn ứng với mỗi trạng thái của tất cả các mã cổ phiếu.

Tại tầng 1 của **Hình 4.1**, thực hiện mờ hóa bộ dữ liệu thực nghiệm tại **Bảng 4.1** bằng cách dựa vào kết quả của **Bảng 4.2** và hàm Gaussian thu được kết quả mờ hóa như **Bảng 4.3**.

Bảng 4.3. Kết quả mờ hóa

X1				X2				X3				X4				X5			
O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
0.593	0.625	0.618	0.662	0.655	0.643	0.629	0.627	0.643	0.644	0.639	0.628	0.621	0.585	0.606	0.596	0.564	0.617	0.568	0.617
0.993	0.979	0.994	0.984	0.960	0.958	0.970	0.963	0.970	0.979	0.969	0.968	0.962	0.975	0.972	0.976	0.978	0.968	0.967	0.948
0.409	0.439	0.408	0.417	0.421	0.433	0.407	0.418	0.410	0.433	0.405	0.418	0.422	0.445	0.429	0.454	0.446	0.476	0.461	0.510
0.997	1.000	0.998	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000	1.000	0.997	0.999	0.996

0.239	0.215	0.237	0.214	0.208	0.219	0.213	0.222	0.224	0.210	0.211	0.215	0.208	0.203	0.212	0.200	0.209	0.202	0.210	0.199
0.997	1.000	0.998	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000	1.000	0.997	0.999	0.996
0.247	0.222	0.246	0.229	0.231	0.234	0.228	0.229	0.231	0.225	0.226	0.230	0.223	0.226	0.236	0.222	0.232	0.216	0.226	0.214
0.498	0.554	0.522	0.555	0.571	0.501	0.545	0.496	0.477	0.490	0.472	0.474	0.502	0.491	0.475	0.512	0.458	0.487	0.461	0.486
0.327	0.297	0.316	0.297	0.318	0.320	0.334	0.325	0.348	0.329	0.352	0.355	0.368	0.360	0.386	0.357	0.402	0.377	0.405	0.376
0.453	0.428	0.452	0.439	0.454	0.478	0.474	0.485	0.512	0.501	0.460	0.463	0.478	0.468	0.441	0.466	0.516	0.487	0.520	0.486
0.114	0.100	0.107	0.103	0.100	0.102	0.102	0.103	0.104	0.096	0.101	0.108	0.095	0.097	0.101	0.090	0.095	0.092	0.096	0.090
0.181	0.155	0.166	0.154	0.161	0.170	0.165	0.173	0.168	0.163	0.163	0.167	0.143	0.152	0.152	0.143	0.149	0.145	0.145	0.142
0.734	0.763	0.736	0.766	0.770	0.747	0.758	0.755	0.748	0.771	0.779	0.788	0.795	0.804	0.783	0.795	0.799	0.820	0.792	0.823
0.629	0.577	0.606	0.591	0.595	0.607	0.592	0.579	0.583	0.596	0.591	0.616	0.597	0.609	0.606	0.584	0.601	0.581	0.580	0.545
0.065	0.077	0.066	0.076	0.073	0.066	0.066	0.070	0.067	0.071	0.064	0.064	0.066	0.071	0.062	0.069	0.066	0.059	0.064	0.060
0.264	0.238	0.254	0.229	0.231	0.250	0.244	0.245	0.256	0.240	0.234	0.238	0.223	0.226	0.228	0.215	0.216	0.216	0.218	0.199
0.357	0.406	0.376	0.396	0.399	0.390	0.385	0.375	0.378	0.379	0.383	0.366	0.338	0.350	0.344	0.347	0.330	0.367	0.342	0.365
0.162	0.167	0.173	0.166	0.155	0.164	0.172	0.173	0.156	0.169	0.197	0.167	0.208	0.183	0.198	0.215	0.202	0.209	0.203	0.221
0.569	0.625	0.570	0.615	0.607	0.607	0.617	0.615	0.631	0.644	0.663	0.663	0.717	0.704	0.726	0.728	0.709	0.700	0.713	0.735
0.957	0.975	0.965	0.972	0.974	0.952	0.965	0.945	0.942	0.954	0.947	0.952	0.962	0.965	0.956	0.972	0.964	0.973	0.967	0.970

Các kết quả mờ hóa trong **Bảng 4.3** được tính bằng công thức (4.1). Cụ thể:

- Mờ hóa giá mở <Open> tại ngày thứ nhất của mã cổ phiếu AAA.

$$\mu(x_{11}^{(1)}) = \exp\left[-\left(\frac{x_{11}^{(1)} - c_1^{(1)}}{a_1^{(1)}}\right)^2\right] = \exp\left[-\left(\frac{16.7 - 11.5}{7.2}\right)^2\right] = 0.593$$

- Mờ hóa giá cao nhất <High> tại ngày thứ nhất của mã cổ phiếu AAA.

$$\mu(x_{12}^{(1)}) = \exp\left[-\left(\frac{x_{12}^{(1)} - c_2^{(1)}}{a_2^{(1)}}\right)^2\right] = \exp\left[-\left(\frac{16.9 - 12.0}{7.2}\right)^2\right] = 0.625$$

- Mờ hóa giá thấp nhất <Low> tại ngày thứ nhất của mã cổ phiếu AAA.

$$\mu(x_{13}^{(1)}) = \exp\left[-\left(\frac{x_{13}^{(1)} - c_3^{(1)}}{a_3^{(1)}}\right)^2\right] = \exp\left[-\left(\frac{16.4 - 11.5}{7.1}\right)^2\right] = 0.618$$

- Mờ hóa giá đóng <Close> tại ngày thứ nhất của mã cổ phiếu AAA.

$$\mu(x_{14}^{(1)}) = \exp\left[-\left(\frac{x_{14}^{(1)} - c_4^{(1)}}{a_4^{(1)}}\right)^2\right] = \exp\left[-\left(\frac{16.4 - 11.8}{7.2}\right)^2\right] = 0.662$$

- Thực hiện tương tự cho các mã cổ phiếu còn lại với các ngày tương ứng sẽ thu được kết quả như **Bảng 4.3**.

Tại tầng 2 của **Hình 4.1**, thực hiện tính tích trọng số mờ từ các giá trị mờ trong **Bảng 4.3** sẽ thu được kết quả như **Bảng 4.4**.

Bảng 4.4. Các giá trị trọng số mờ

W1				W2				W3				W4				W5			
O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
5.3E-09	4.7E-09	5.2E-09	4.8E-09	4.8E-09	5E-09	5E-09	5.1E-09	5.2E-09	5E-09	5E-09	5.1E-09	4.8E-09	4.9E-09	5E-09	4.7E-09	5E-09	4.7E-09	4.9E-09	4.5E-09

Các giá trị trọng số mờ trong **Bảng 4.4** tham gia quá trình tổng hợp suy luận để tìm ra kết quả. Các giá trị này được tính toán như sau:

- Trọng số mờ giá mở <Open> tại ngày thứ nhất.

$$W_1^{(1)} = 0.593 \times 0.993 \times 0.409 \times 0.997 \times 0.239 \times 0.997 \times 0.247 \times 0.498 \times 0.327 \times 0.453 \times 0.114 \times 0.181 \times 0.734 \times 0.629 \times 0.065 \times 0.264 \times 0.357 \times 0.162 \times 0.569 \times 0.957 = 5.3E - 09$$

- Trọng số mờ giá cao nhất <High> tại ngày thứ nhất.

$$W_2^{(1)} = 0.625 \times 0.979 \times 0.439 \times 1 \times 0.215 \times 1 \times 0.222 \times 0.554 \times 0.297 \times 0.428 \times 0.1 \times 0.155 \times 0.763 \times 0.577 \times 0.077 \times 0.238 \times 0.406 \times 0.167 \times 0.625 \times 0.975 = 4.7E - 09$$

- Trọng số mờ giá thấp nhất <Low> tại ngày thứ nhất.

$$W_3^{(1)} = 0.618 \times 0.994 \times 0.408 \times 0.998 \times 0.237 \times 0.998 \times 0.246 \times 0.522 \times 0.316 \times 0.452 \times 0.107 \times 0.166 \times 0.736 \times 0.606 \times 0.066 \times 0.254 \times 0.376 \times 0.173 \times 0.57 \times 0.965 = 5.2E - 09$$

- Trọng số mờ giá đóng <Close> tại ngày thứ nhất.

$$W_4^{(1)} = 0.662 \times 0.984 \times 0.417 \times 1 \times 0.214 \times 1 \times 0.229 \times 0.555 \times 0.297 \times 0.439 \times 0.103 \times 0.154 \times 0.766 \times 0.591 \times 0.076 \times 0.229 \times 0.396 \times 0.166 \times 0.615 \times 0.972 = 4.8E - 09$$

- Thực hiện tương tự cho các ngày còn lại thu được kết quả như **Bảng 4.4**.

Tại tầng 3 **Hình 4.1**, chuẩn hóa các giá trị trọng số mờ **Bảng 4.4** được **Bảng 4.5**.

Bảng 4.5. Chuẩn hóa các giá trị trọng số mờ

$\overline{W1}$				$\overline{W2}$				$\overline{W3}$				$\overline{W4}$				$\overline{W5}$			
O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
0.213	0.194	0.207	0.199	0.191	0.207	0.199	0.210	0.206	0.205	0.198	0.211	0.191	0.200	0.198	0.194	0.200	0.193	0.197	0.186

Các giá trị trọng số mờ được chuẩn hóa trong **Bảng 4.5** giúp tìm ra mức độ tham gia của trọng số trong quá trình tổng hợp suy luận kết quả. Việc chuẩn hóa các giá trị trọng số mờ được tính toán như sau:

- Chuẩn hóa trọng số mờ giá mở <Open> tại ngày thứ nhất.

$$\begin{aligned}\overline{W_1^{(1)}} &= \frac{W_1^{(1)}}{W_1^{(1)} + W_1^{(2)} + W_1^{(3)} + W_1^{(4)} + W_1^{(5)}} \\ &= \frac{5.3E - 0.9}{5.3E - 0.9 + 4.8E - 0.9 + 5.2E - 0.9 + 4.8E - 0.9 + 5.0E - 0.9} = 0.213\end{aligned}$$

- Chuẩn hóa trọng số mờ giá cao nhất <High> tại ngày thứ nhất.

$$\begin{aligned}\overline{W_2^{(1)}} &= \frac{W_2^{(1)}}{W_2^{(1)} + W_2^{(2)} + W_2^{(3)} + W_2^{(4)} + W_2^{(5)}} \\ &= \frac{4.7E - 0.9}{4.7E - 0.9 + 5.0E - 0.9 + 5.0E - 0.9 + 4.9E - 0.9 + 4.7E - 0.9} = 0.194\end{aligned}$$

- Chuẩn hóa trọng số mờ giá thấp nhất <Low> tại ngày thứ nhất.

$$\begin{aligned}\overline{W_3^{(1)}} &= \frac{W_3^{(1)}}{W_3^{(1)} + W_3^{(2)} + W_3^{(3)} + W_3^{(4)} + W_3^{(5)}} \\ &= \frac{5.2E - 0.9}{5.2E - 0.9 + 5.0E - 0.9 + 5.0E - 0.9 + 5.0E - 0.9 + 4.9E - 0.9} = 0.207\end{aligned}$$

- Chuẩn hóa trọng số mờ giá đóng <Close> tại ngày thứ nhất.

$$\begin{aligned}\overline{W_4^{(1)}} &= \frac{W_4^{(1)}}{W_4^{(1)} + W_4^{(2)} + W_4^{(3)} + W_4^{(4)} + W_4^{(5)}} \\ &= \frac{4.8E - 0.9}{4.8E - 0.9 + 5.1E - 0.9 + 5.1E - 0.9 + 4.7E - 0.9 + 4.5E - 0.9} = 0.199\end{aligned}$$

- Thực hiện chuẩn hóa tương tự cho các ngày còn lại sẽ thu được kết quả như **Bảng 4.5**.

Tại tầng 4 của **Hình 4.1**, phát sinh ngẫu nhiên ma trận hệ số suy luận $R = (r_{ij})$ ứng với các trạng thái <Open>, <High>, <Low>, <Close> bằng phương pháp suy luận mờ Takagi - Sugeno. Kết quả như **Bảng 4.6**.

Bảng 4.6. Ma trận hệ số suy luận mờ lần huấn luyện thứ 1

<Open>						<High>					
R0	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
0.183	0.098	0.202	0.107	0.200	0.154	0.241	0.045	0.035	0.117	0.049	0.145
0.020	0.066	0.113	0.046	0.180	0.004	0.241	0.144	0.049	0.196	0.236	0.103
0.128	0.051	0.224	0.030	0.117	0.156	0.100	0.122	0.061	0.078	0.029	0.109
0.060	0.223	0.157	0.125	0.222	0.190	0.150	0.140	0.205	0.009	0.011	0.215
0.223	0.011	0.148	0.070	0.074	0.178	0.206	0.110	0.175	0.185	0.188	0.125
<Low>						<Close>					
R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	R23
0.092	0.073	0.201	0.107	0.084	0.187	0.132	0.043	0.012	0.096	0.119	0.162
0.145	0.010	0.082	0.158	0.192	0.051	0.149	0.085	0.002	0.033	0.240	0.028
0.021	0.046	0.018	0.141	0.141	0.068	0.194	0.249	0.197	0.150	0.133	0.142
0.199	0.121	0.153	0.218	0.087	0.234	0.129	0.129	0.133	0.002	0.188	0.107
0.144	0.037	0.185	0.020	0.128	0.091	0.062	0.121	0.086	0.173	0.030	0.247

Các giá trị của ma trận hệ số suy luận mờ trong **Bảng 4.6** được phát sinh ngẫu nhiên. Số dòng, số cột của ma trận được xác định dựa vào công thức (4.10). Với dữ liệu thực nghiệm gồm 5 ngày ($N = 5$) và 4 trạng thái ($n = 4$) thì ma trận hệ số suy luận mờ sẽ bao gồm 5 dòng và 24 cột.

Thực hiện tính các giá trị dự báo $f_i^{(N)}$ tại tầng 4 của **Hình 4.1** dựa vào ma trận hệ số suy luận mờ trong **Bảng 4.6** và dữ liệu thực nghiệm tại **Bảng 4.1** sẽ thu được kết quả như **Bảng 4.7**.

Bảng 4.7. Các giá trị dự báo $f_i^{(N)}$ lần huấn luyện thứ 1

$f1$				$f2$				$f3$				$f4$				$f5$			
O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
12.83	6.83	10.89	7.37	6.78	12.54	8.24	6.64	9.75	6.80	6.83	14.66	15.34	9.91	13.64	9.47	8.25	13.39	7.77	11.01
8.14	4.43	6.86	4.65	4.28	8.07	5.22	4.28	6.17	4.36	4.30	9.36	9.72	6.30	8.65	6.04	5.25	8.56	4.92	6.91
14.09	7.41	11.91	7.94	7.52	13.62	9.07	7.19	10.69	7.41	7.50	16.09	16.83	10.79	14.92	10.30	9.01	14.59	8.51	11.99
8.94	4.81	7.51	5.12	4.72	8.80	5.75	4.67	6.80	4.76	4.72	10.29	10.56	6.93	9.43	6.63	5.79	9.39	5.40	7.64
2.33	1.37	1.89	1.35	1.18	2.35	1.48	1.24	1.76	1.26	1.15	2.67	2.67	1.84	2.44	1.73	1.58	2.47	1.41	1.93
9.19	4.99	7.75	5.34	4.87	9.01	5.89	4.76	6.97	4.92	4.85	10.58	10.94	7.19	9.74	6.82	5.92	9.63	5.55	7.97
2.50	1.45	2.02	1.45	1.26	2.50	1.60	1.34	1.90	1.33	1.24	2.84	2.85	1.95	2.60	1.85	1.70	2.63	1.51	2.06
13.53	7.25	11.50	7.80	7.16	13.24	8.79	6.96	10.25	7.19	7.29	15.54	16.20	10.50	14.45	9.98	8.69	14.19	8.18	11.73
3.53	2.00	2.96	2.10	1.79	3.45	2.33	1.87	2.68	1.84	1.85	3.92	4.07	2.69	3.77	2.57	2.39	3.64	2.17	2.98
4.42	2.50	3.65	2.56	2.26	4.37	2.75	2.29	3.35	2.35	2.22	4.99	5.14	3.42	4.62	3.25	2.95	4.65	2.63	3.73
0.83	0.60	0.61	0.51	0.37	0.91	0.54	0.48	0.62	0.47	0.35	0.99	0.85	0.69	0.85	0.63	0.63	0.92	0.51	0.66
1.69	1.06	1.34	0.99	0.82	1.76	1.08	0.90	1.28	0.94	0.80	1.96	1.88	1.37	1.75	1.25	1.18	1.86	1.02	1.39
11.80	6.26	9.98	6.69	6.27	11.47	7.57	6.05	8.95	6.26	6.26	13.53	14.09	9.14	12.51	8.69	7.56	12.31	7.14	10.07
5.18	2.86	4.28	2.95	2.71	5.13	3.31	2.70	3.94	2.76	2.68	5.87	6.09	4.02	5.41	3.77	3.39	5.46	3.11	4.33
17.96	9.52	15.30	10.32	9.58	17.44	11.61	9.22	13.63	9.53	9.64	20.62	21.51	13.95	19.13	13.24	11.46	18.73	10.90	15.56
2.52	1.49	2.05	1.42	1.28	2.58	1.61	1.32	1.89	1.39	1.25	2.86	2.89	2.02	2.63	1.83	1.69	2.73	1.53	2.04
14.55	7.68	12.27	8.33	7.74	14.10	9.33	7.50	11.04	7.64	7.73	16.62	17.40	11.15	15.36	10.71	9.32	15.08	8.76	12.48

16.19	8.55	13.62	9.14	8.62	15.78	10.29	8.22	12.30	8.59	8.53	18.55	19.35	12.52	17.01	11.83	10.34	16.92	9.72	13.87
12.59	6.69	10.61	7.06	6.70	12.25	7.97	6.37	9.57	6.71	6.60	14.45	15.05	9.81	13.29	9.22	8.04	13.16	7.56	10.73
10.12	5.43	8.55	5.80	5.36	9.90	6.56	5.20	7.65	5.37	5.40	11.65	12.05	7.83	10.75	7.47	6.51	10.61	6.12	8.72

Thực hiện tính giá trị dự báo $f_i^{(N)}$ bằng cách dựa vào các dữ liệu thực nghiệm và công thức (4.10). Cụ thể:

- Giá trị dự báo mã cổ phiếu AAA với trạng thái <Open> tại ngày thứ nhất.

$$f_1^{(1)} = 0.183 + 0.098 \times 16.7 + 0.202 \times 16.4 + 0.107 \times 16.4 + 0.2 \times 16.6 + 0.154 \times 17 = 12.83$$

- Giá trị dự báo mã cổ phiếu AAA với trạng thái <High> tại ngày thứ nhất.

$$f_2^{(1)} = 0.241 + 0.045 \times 16.9 + 0.035 \times 16.6 + 0.117 \times 16.6 + 0.049 \times 17.1 + 0.145 \times 17 = 6.83$$

- Giá trị dự báo mã cổ phiếu AAA với trạng thái <Low> tại ngày thứ nhất.

$$f_3^{(1)} = 0.092 + 0.073 \times 16.4 + 0.201 \times 16.4 + 0.107 \times 16.3 + 0.084 \times 16.5 + 0.187 \times 16.9 = 10.89$$

- Giá trị dự báo mã cổ phiếu AAA với trạng thái <Close> tại ngày thứ nhất.

$$f_4^{(1)} = 0.132 + 0.043 \times 16.4 + 0.012 \times 16.6 + 0.096 \times 16.5 + 0.119 \times 16.9 + 0.162 \times 16.9 = 7.37$$

- Thực hiện tương tự cho các ngày còn lại ứng với từng mã cổ phiếu sẽ thu được kết quả như **Bảng 4.7**.

Tại tầng 5 của **Hình 4.1**, thực hiện chuẩn hóa ma trận dự báo dựa vào các giá trị trọng số mờ được chuẩn hóa tại **Bảng 4.5** và các giá trị dự báo $f_i^{(N)}$ tại **Bảng 4.7**. Kết quả thu được ma trận dự báo như **Bảng 4.8**.

Bảng 4.8. Chuẩn hóa các giá trị dự báo lần huấn luyện thứ 1

$\overline{W1} \times f1$				$\overline{W2} \times f2$				$\overline{W3} \times f3$				$\overline{W4} \times f4$				$\overline{W5} \times f5$			
O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
2.726	1.328	2.259	1.468	1.293	2.600	1.642	1.396	2.006	1.392	1.353	3.090	2.933	1.985	2.699	1.836	1.649	2.586	1.534	2.044
1.730	0.862	1.423	0.927	0.815	1.672	1.039	0.901	1.269	0.894	0.853	1.973	1.858	1.263	1.711	1.171	1.050	1.654	0.971	1.284
2.996	1.440	2.472	1.582	1.433	2.823	1.805	1.512	2.201	1.517	1.486	3.392	3.217	2.162	2.954	1.998	1.800	2.818	1.679	2.226
1.899	0.935	1.559	1.021	0.899	1.823	1.145	0.982	1.401	0.974	0.934	2.169	2.019	1.388	1.867	1.286	1.157	1.814	1.065	1.419
0.496	0.267	0.392	0.270	0.224	0.486	0.295	0.261	0.361	0.259	0.228	0.564	0.510	0.370	0.483	0.336	0.316	0.477	0.278	0.358
1.954	0.970	1.608	1.064	0.928	1.868	1.173	1.001	1.435	1.007	0.961	2.230	2.091	1.440	1.928	1.323	1.182	1.861	1.095	1.481
0.532	0.283	0.419	0.289	0.240	0.518	0.318	0.282	0.391	0.273	0.247	0.599	0.545	0.390	0.515	0.358	0.340	0.509	0.298	0.382

2.877	1.409	2.386	1.555	1.365	2.745	1.750	1.463	2.109	1.472	1.445	3.276	3.097	2.103	2.861	1.936	1.738	2.742	1.615	2.178
0.750	0.390	0.614	0.419	0.340	0.716	0.464	0.393	0.551	0.376	0.367	0.827	0.778	0.539	0.746	0.499	0.478	0.703	0.428	0.553
0.940	0.485	0.758	0.510	0.432	0.906	0.548	0.482	0.689	0.481	0.439	1.052	0.983	0.686	0.915	0.631	0.590	0.899	0.519	0.692
0.177	0.116	0.127	0.102	0.070	0.189	0.107	0.101	0.128	0.096	0.070	0.208	0.162	0.137	0.168	0.121	0.126	0.178	0.101	0.123
0.359	0.206	0.277	0.196	0.157	0.365	0.215	0.189	0.263	0.192	0.159	0.413	0.360	0.274	0.346	0.243	0.235	0.359	0.201	0.258
2.509	1.217	2.072	1.332	1.196	2.378	1.507	1.273	1.842	1.282	1.239	2.851	2.693	1.831	2.476	1.686	1.510	2.379	1.410	1.870
1.102	0.556	0.888	0.587	0.516	1.064	0.660	0.567	0.811	0.564	0.531	1.237	1.165	0.806	1.071	0.731	0.677	1.055	0.614	0.804
3.817	1.852	3.174	2.057	1.825	3.615	2.312	1.938	2.805	1.950	1.910	4.346	4.111	2.794	3.787	2.569	2.291	3.618	2.151	2.890
0.536	0.290	0.426	0.283	0.245	0.534	0.320	0.277	0.389	0.284	0.247	0.602	0.553	0.404	0.521	0.356	0.337	0.527	0.302	0.378
3.092	1.494	2.546	1.660	1.474	2.922	1.857	1.577	2.272	1.565	1.532	3.502	3.326	2.233	3.040	2.077	1.863	2.913	1.729	2.318
3.441	1.663	2.825	1.822	1.643	3.271	2.050	1.728	2.532	1.760	1.690	3.909	3.698	2.508	3.368	2.295	2.067	3.269	1.918	2.576
2.676	1.300	2.201	1.407	1.277	2.539	1.587	1.339	1.970	1.374	1.308	3.045	2.878	1.965	2.630	1.788	1.608	2.543	1.493	1.993
2.151	1.056	1.773	1.156	1.022	2.051	1.306	1.093	1.574	1.099	1.070	2.454	2.304	1.569	2.128	1.449	1.300	2.049	1.207	1.619

Thực hiện chuẩn hóa ma trận dự báo bằng công thức $\overline{W_n^{(N)}} \times f_n^{(N)}$. Cụ thể:

- Chuẩn hóa giá trị dự báo mã cổ phiếu AAA ứng với trạng thái <Open> tại ngày thứ nhất.

$$\overline{W_1^{(1)}} \times f_1^{(1)} = 0.213 \times 12.83 = 2.726$$

- Chuẩn hóa giá trị dự báo mã cổ phiếu AAA ứng với trạng thái <High> tại ngày thứ nhất.

$$\overline{W_2^{(1)}} \times f_2^{(1)} = 0.194 \times 6.83 = 1.328$$

- Chuẩn hóa giá trị dự báo mã cổ phiếu AAA ứng với trạng thái <Low> tại ngày thứ nhất.

$$\overline{W_3^{(1)}} \times f_3^{(1)} = 0.207 \times 10.89 = 2.259$$

- Chuẩn hóa giá trị dự báo mã cổ phiếu AAA ứng với trạng thái <Close> tại ngày thứ nhất.

$$\overline{W_4^{(1)}} \times f_4^{(1)} = 0.199 \times 7.37 = 1.468$$

- Thực hiện chuẩn hóa tương tự cho các ngày còn lại ứng với từng mã cổ phiếu sẽ thu được kết quả như **Bảng 4.8**.

Tổng hợp kết quả từ ma trận giá trị dự báo đã được chuẩn hóa sẽ tìm ra giá trị dự báo, độ lệch. Kết quả như **Bảng 4.9**.

Bảng 4.9. Kết quả dự báo lần huấn luyện thứ 1

Giá trị dự báo				Giá trị thực tế (X6)				Độ lệch			
O	H	L	C	O	H	L	C	O	H	L	C
10.607	9.891	9.487	9.834	17	17	16.8	16.9	6.393	7.109	7.313	7.066
6.722	6.344	5.998	6.256	17	17	16.8	16.9	10.278	10.656	10.802	10.644
11.647	10.760	10.395	10.710	18.3	18.3	17.8	17.9	6.653	7.540	7.405	7.190
7.375	6.935	6.570	6.878	11.5	11.7	11.4	11.7	4.125	4.765	4.830	4.822
1.907	1.859	1.676	1.788	3.1	3.1	3	3	1.193	1.241	1.324	1.212
7.590	7.146	6.764	7.098	12.5	12.6	12.2	12.4	4.910	5.454	5.436	5.302
2.047	1.973	1.796	1.910	3	3	2.9	3	0.953	1.027	1.104	1.090
11.185	10.471	10.056	10.408	17.9	18.3	17.9	18.2	6.715	7.829	7.844	7.792
2.897	2.724	2.619	2.691	5.2	5.2	5.1	5.2	2.303	2.476	2.481	2.509
3.634	3.456	3.179	3.366	6.3	6.3	6.3	6.3	2.666	2.844	3.121	2.934
0.663	0.717	0.574	0.656	0.8	0.9	0.8	0.9	0.137	0.183	0.226	0.244
1.374	1.395	1.198	1.300	1.9	2	1.9	2	0.526	0.605	0.702	0.700
9.749	9.087	8.704	9.012	15.1	15.5	15	15	5.351	6.413	6.296	5.988
4.271	4.046	3.763		6.5	6.5	6.4	6.4	2.229	2.454	2.637	2.473
14.849	13.829	13.333	13.801	24	24.7	23.9	24.3	9.151	10.871	10.567	10.499
2.060	2.040	1.815	1.895	3	3.1	2.9	3	0.940	1.060	1.085	1.105
12.026	11.126	10.704	11.134	19.1	19.4	18.9	19.2	7.074	8.274	8.196	8.066
13.381	12.472	11.850	12.330	21	21	20.6	20.6	7.619	8.528	8.750	8.270
10.408	9.722	9.219	9.573	16.2	16.2	15.9	16	5.792	6.478	6.681	6.427
8.351	7.824	7.484	7.771	13.1	13.2	13	13.1	4.749	5.376	5.516	5.329
Trung bình độ lệch								4.488	5.059	5.116	4.983

Dựa vào **Bảng 4.8** để tính toán giá trị dự báo cho từng mã cổ phiếu, tính đoán độ lệch dựa vào giá trị thực tế và giá trị thực tế. Cụ thể:

- Giá trị dự báo, độ lệch mã cổ phiếu AAA ứng với trạng thái <Open>.

$$F_{11} = \overline{W}_1^{(1)} f_1^{(1)} + \overline{W}_1^{(2)} f_1^{(2)} + \overline{W}_1^{(3)} f_1^{(3)} + \overline{W}_1^{(4)} f_1^{(4)} + \overline{W}_1^{(5)} f_1^{(5)}$$

$$= 2.726 + 1.293 + 2.006 + 2.933 + 1.649 = 10.607$$

$$\Delta_{11} = F_{11}^{(6)} - F_{11} = 17 - 10.607 = 6.393$$

- Giá trị dự báo, độ lệch mã cổ phiếu AAA ứng với trạng thái <High>.

$$F_{12} = \overline{W}_2^{(1)} f_2^{(1)} + \overline{W}_2^{(2)} f_2^{(2)} + \overline{W}_2^{(3)} f_2^{(3)} + \overline{W}_2^{(4)} f_2^{(4)} + \overline{W}_2^{(5)} f_2^{(5)}$$

$$= 1.328 + 2.6 + 1.392 + 1.985 + 2.586 = 9.891$$

$$\Delta_{12} = F_{12}^{(6)} - F_{12} = 17 - 9.891 = 7.109$$

- Giá trị dự báo, độ lệch mã cổ phiếu AAA ứng với trạng thái <Low>.

$$F_{13} = \overline{W}_3^{(1)} f_3^{(1)} + \overline{W}_3^{(2)} f_3^{(2)} + \overline{W}_3^{(3)} f_3^{(3)} + \overline{W}_3^{(4)} f_3^{(4)} + \overline{W}_3^{(5)} f_3^{(5)}$$

$$= 2.259 + 1.642 + 1.353 + 2.699 + 1.534 = 9.487$$

$$\Delta_{13} = F_{13}^{(6)} - F_{13} = 16.8 - 9.487 = 7.313$$

- Giá trị dự báo, độ lệch mã cổ phiếu AAA ứng với trạng thái <Close>.

$$\begin{aligned} F_{14} &= \overline{W}_4^{(1)} f_4^{(1)} + \overline{W}_4^{(2)} f_4^{(2)} + \overline{W}_4^{(3)} f_4^{(3)} + \overline{W}_4^{(4)} f_4^{(4)} + \overline{W}_4^{(5)} f_4^{(5)} \\ &= 1.468 + 1.396 + 3.09 + 1.836 + 2.044 = 9.83 \end{aligned}$$

$$\Delta_{14} = F_{14}^{(6)} - F_{14} = 16.9 - 9.834 = 7.066$$

- Thực hiện tương tự cho các mã cổ phiếu còn lại sẽ thu được kết quả như

Bảng 4.9.

Như vậy, trung bình độ lệch giữa giá trị thực tế và giá trị dự báo ở lần huấn luyện thứ nhất ứng với trạng thái <Open> là 4.488, <High> là 5.059, <Low> là 5.116, <Close> là 4.983 (đơn vị tính 1,000 VNĐ).

Thực hiện huấn luyện tương tự như vậy cho các lần huấn luyện tiếp theo. Sau một quá trình huấn luyện thu được ma trận hệ số suy luận $R = (r_{ij})$ như sau:

Bảng 4.10. Ma trận hệ số suy luận mờ lần huấn luyện thứ 8

<Open>						<High>					
R0	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
0.022	0.114	0.534	0.034	0.098	0.479	0.061	0.094	0.135	0.063	0.085	0.216
0.036	0.028	0.024	0.088	0.295	0.373	0.036	0.416	0.022	0.357	0.733	0.031
0.119	0.452	0.008	0.207	0.628	0.070	0.090	0.569	0.052	0.090	0.512	0.067
0.305	0.131	0.084	0.138	0.032	0.116	0.041	0.045	0.538	0.029	0.330	0.030
0.043	0.001	0.044	0.048	0.016	0.869	0.123	0.154	0.350	0.003	0.042	0.078
<Low>						<Close>					
R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	R23
0.104	0.002	0.091	0.100	0.065	0.326	0.018	0.087	0.057	0.461	0.077	0.059
0.000	0.401	0.187	0.099	0.613	0.078	0.065	0.163	0.091	0.281	0.968	0.366
0.037	0.140	0.564	0.117	0.176	0.315	0.102	0.056	0.092	0.140	0.058	0.174
0.068	0.133	0.007	0.199	0.084	0.554	0.121	0.049	0.342	0.042	0.458	0.045
0.600	0.102	0.181	0.259	0.122	0.030	0.005	0.002	0.166	0.073	0.370	0.266

Bảng 4.11. Các giá trị dự báo $f_i^{(N)}$ lần huấn luyện thứ 8

f1				f2				f3				f4				f5			
O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
21.01	10.06	9.82	12.29	13.57	26.40	22.68	31.43	22.81	21.95	21.72	8.77	8.62	16.36	16.36	15.78	16.59	10.62	11.98	14.74
13.21	6.38	6.11	7.73	8.49	16.82	14.43	19.80	14.49	13.99	13.64	5.51	5.58	10.23	10.21	9.97	10.37	6.72	7.81	9.18
22.98	10.93	10.58	13.55	14.72	28.67	24.95	33.93	25.10	23.81	23.83	9.53	9.44	17.89	17.64	17.17	17.67	11.65	13.23	15.86
14.56	7.00	6.74	8.53	9.32	18.38	15.58	21.77	15.64	15.28	14.95	6.09	6.02	11.44	11.17	10.98	11.29	7.46	8.48	10.15
3.56	1.78	1.71	2.07	2.31	4.52	3.81	5.33	4.01	3.84	3.64	1.57	1.73	2.83	2.79	2.78	2.78	1.95	2.50	2.50
14.90	7.25	6.95	8.81	9.60	18.71	16.10	22.44	16.34	15.59	15.42	6.35	6.22	11.63	11.56	11.26	11.57	7.65	8.72	10.56

3.90	1.90	1.83	2.22	2.50	4.87	4.10	5.79	4.22	4.09	3.91	1.67	1.82	3.05	2.98	2.98	3.06	2.07	2.64	2.69
22.05	10.67	10.39	13.15	14.37	27.76	24.02	33.12	24.16	22.95	22.87	9.33	9.11	17.37	17.34	16.63	17.49	11.26	12.71	15.57
5.60	2.71	2.80	3.22	3.81	6.84	5.92	8.53	6.04	5.65	5.74	2.42	2.46	4.22	4.56	4.22	4.76	2.76	3.55	4.01
7.06	3.44	3.38	4.02	4.67	8.75	7.16	10.51	7.60	7.19	7.17	3.00	3.08	5.49	5.54	5.34	5.77	3.61	4.27	4.97
1.10	0.60	0.57	0.69	0.70	1.48	1.10	1.67	1.28	1.31	1.09	0.57	0.74	0.92	0.85	0.92	0.83	0.70	1.16	0.76
2.54	1.30	1.19	1.50	1.57	3.27	2.58	3.72	2.78	2.75	2.55	1.15	1.32	2.10	1.88	2.01	1.92	1.46	1.95	1.75
19.21	9.21	8.88	11.27	12.25	24.04	20.89	28.48	20.99	20.03	19.96	8.01	7.97	15.11	14.80	14.44	14.74	9.90	11.13	13.35
8.32	4.03	3.84	4.88	5.35	10.51	8.88	12.31	9.09	8.72	8.45	3.48	3.58	6.58	6.33	6.22	6.48	4.31	5.05	5.71
29.38	14.15	13.68	17.42	18.90	36.67	32.06	43.99	32.05	30.44	30.60	12.35	11.98	23.00	22.83	22.03	22.81	14.96	16.75	20.66
3.84	1.96	1.82	2.25	2.46	5.02	4.12	5.67	4.36	4.23	3.99	1.67	1.87	3.18	2.95	3.00	2.91	2.17	2.69	2.64
23.76	11.34	11.03	13.98	15.41	29.67	25.70	35.58	25.97	24.57	24.49	9.93	9.71	18.46	18.39	17.85	18.66	11.95	13.48	16.68
26.54	12.68	12.10	15.66	16.78	33.31	28.50	38.95	28.68	27.68	27.29	11.00	10.86	20.82	20.18	19.66	20.25	13.56	14.98	18.24
20.60	9.87	9.35	12.05	12.91	25.66	22.13	29.99	22.34	21.41	21.26	8.52	8.52	16.17	15.63	15.28	15.56	10.63	11.83	14.06
16.36	7.92	7.66	9.77	10.60	20.64	17.82	24.53	18.01	17.08	17.00	6.95	6.85	12.98	12.73	12.43	12.76	8.45	9.63	11.54

Bảng 4.12. Chuẩn hóa các giá trị dự báo lần huấn luyện thứ 8

$\overline{W1} \times f1$				$\overline{W2} \times f2$				$\overline{W3} \times f3$				$\overline{W4} \times f4$				$\overline{W5} \times f5$			
O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C	O	H	L	C
4.466	1.957	2.037	2.449	2.587	5.473	4.517	6.609	4.695	4.494	4.303	1.847	1.647	3.277	3.238	3.062	3.316	2.052	2.364	2.738
2.807	1.240	1.267	1.540	1.619	3.486	2.873	4.163	2.983	2.865	2.702	1.160	1.067	2.050	2.022	1.933	2.073	1.298	1.540	1.706
4.884	2.125	2.196	2.700	2.805	5.942	4.968	7.135	5.166	4.875	4.720	2.007	1.804	3.583	3.491	3.330	3.532	2.250	2.611	2.945
3.095	1.360	1.399	1.700	1.776	3.809	3.103	4.578	3.219	3.128	2.963	1.283	1.150	2.292	2.211	2.130	2.256	1.442	1.673	1.886
0.757	0.346	0.355	0.412	0.440	0.937	0.758	1.121	0.825	0.785	0.721	0.332	0.331	0.567	0.552	0.539	0.556	0.377	0.493	0.464
3.167	1.410	1.441	1.755	1.829	3.878	3.207	4.719	3.362	3.192	3.055	1.338	1.189	2.330	2.289	2.185	2.313	1.478	1.720	1.962
0.829	0.369	0.380	0.442	0.476	1.009	0.817	1.218	0.869	0.837	0.774	0.353	0.349	0.612	0.590	0.578	0.613	0.399	0.520	0.500
4.688	2.074	2.155	2.620	2.738	5.755	4.782	6.964	4.972	4.700	4.531	1.966	1.741	3.479	3.433	3.225	3.495	2.175	2.508	2.892
1.190	0.527	0.582	0.641	0.725	1.419	1.178	1.793	1.243	1.156	1.137	0.509	0.471	0.844	0.903	0.818	0.951	0.533	0.701	0.745
1.501	0.668	0.702	0.801	0.890	1.814	1.426	2.209	1.564	1.472	1.421	0.633	0.589	1.099	1.096	1.037	1.154	0.697	0.842	0.923
0.233	0.117	0.118	0.137	0.133	0.307	0.219	0.350	0.263	0.268	0.215	0.120	0.142	0.184	0.168	0.179	0.167	0.136	0.228	0.141
0.540	0.252	0.247	0.299	0.299	0.678	0.513	0.783	0.573	0.564	0.505	0.243	0.253	0.421	0.372	0.391	0.383	0.282	0.384	0.326
4.082	1.790	1.843	2.245	2.334	4.984	4.160	5.989	4.320	4.101	3.954	1.687	1.524	3.027	2.930	2.802	2.946	1.912	2.196	2.479
1.770	0.784	0.796	0.973	1.019	2.178	1.768	2.588	1.870	1.786	1.675	0.732	0.684	1.318	1.253	1.206	1.295	0.833	0.997	1.059
6.246	2.752	2.838	3.471	3.601	7.602	6.383	9.250	6.597	6.234	6.062	2.602	2.290	4.608	4.518	4.273	4.560	2.891	3.306	3.837
0.816	0.381	0.378	0.449	0.468	1.041	0.821	1.192	0.898	0.866	0.791	0.351	0.357	0.636	0.584	0.581	0.581	0.419	0.532	0.491
5.050	2.204	2.289	2.785	2.936	6.150	5.118	7.482	5.345	5.032	4.852	2.092	1.857	3.697	3.641	3.463	3.730	2.309	2.660	3.097
5.640	2.465	2.511	3.121	3.198	6.903	5.676	8.191	5.904	5.667	5.406	2.318	2.077	4.171	3.994	3.814	4.047	2.619	2.956	3.387
4.378	1.918	1.940	2.401	2.461	5.318	4.408	6.306	4.597	4.383	4.211	1.795	1.630	3.240	3.093	2.963	3.110	2.054	2.334	2.612
3.478	1.540	1.588	1.948	2.020	4.277	3.549	5.157	3.707	3.498	3.369	1.465	1.310	2.599	2.520	2.412	2.550	1.632	1.900	2.142

Bảng 4.13. Kết quả dự báo lần huấn luyện thứ 8

Giá trị dự báo				Giá trị thực tế (X6)				Độ lệch			
O	H	L	C	O	H	L	C	O	H	L	C
16.711	17.253	16.459	16.704	17	17	16.8	16.9	0.289	0.253	0.341	0.196
10.548	10.938	10.404	10.503	17	17	16.8	16.9	6.452	6.062	6.396	6.397
18.190	18.775	17.986	18.118	18.3	18.3	17.8	17.9	0.110	0.475	0.186	0.218
11.496	12.033	11.349	11.577	11.5	11.7	11.4	11.7	0.004	0.333	0.051	0.123
2.909	3.013	2.878	2.867	3.1	3.1	3	3	0.191	0.087	0.122	0.133

11.860	12.288	11.711	11.959	12.5	12.6	12.2	12.4	0.640	0.312	0.489	0.441
3.136	3.225	3.081	3.090	3	3	2.9	3	0.136	0.225	0.181	0.090
17.634	18.183	17.409	17.668	17.9	18.3	17.9	18.2	0.266	0.117	0.491	0.532
4.581	4.480	4.500	4.506	5.2	5.2	5.1	5.2	0.619	0.720	0.600	0.694
5.698	5.751	5.488	5.603	6.3	6.3	6.3	6.3	0.602	0.549	0.812	0.697
0.937	1.012	0.949	0.927	0.8	0.9	0.8	0.9	0.137	0.112	0.149	0.027
2.048	2.197	2.022	2.042	1.9	2	1.9	2	0.148	0.197	0.122	0.042
15.207	15.815	15.082	15.203	15.1	15.5	15	15	0.107	0.315	0.082	0.203
6.638	6.899	6.489	6.559	6.5	6.5	6.4	6.4	0.138	0.399	0.089	0.159
23.294	24.085	23.107	23.433	24	24.7	23.9	24.3	0.706	0.615	0.793	0.867
3.119	3.343	3.106	3.064	3	3.1	2.9	3	0.119	0.243	0.206	0.064
18.918	19.391	18.560	18.919	19.1	19.4	18.9	19.2	0.182	0.009	0.340	0.281
20.866	21.825	20.543	20.831	21	21	20.6	20.6	0.134	0.825	0.057	0.231
16.176	16.912	15.986	16.077	16.2	16.2	15.9	16	0.024	0.712	0.086	0.077
13.064	13.547	12.925	13.124	13.1	13.2	13	13.1	0.036	0.347	0.075	0.024
Trung bình độ lệch								0.552	0.645	0.583	0.575

Thực hiện 8 lần huấn luyện, trung bình độ lệch của lần thứ 8 của <Open> là 0.552 (giảm 3.936 so với lần huấn luyện 1), <High> là 0.645 (giảm 4.414 so với lần huấn luyện 1), <Low> là 0.583 (giảm 4.532 so với lần huấn luyện 1), <Close> là 0.575 (giảm 4.408 so với lần huấn luyện 1).

Ví dụ trên đã trình bày từng bước của quá trình suy luận ANFIS đề xuất áp dụng dự báo giá cổ phiếu dựa trên dữ liệu thực tế Cophieu68. Trên cơ sở này luận văn sẽ thực hiện xây dựng ứng dụng thực nghiệm trên bộ dữ liệu lớn hơn nhằm minh chứng tính hiệu quả của phương pháp đề xuất đồng thời so sánh kết quả với các phương pháp đồng dạng khác để từ đó cho thấy tính hiệu quả của phương pháp đề xuất.

4.4. Tổng kết chương

Trong chương này đã tiếp cận mô hình mạng ANFIS truyền thống để từ đó đề xuất mô hình mạng ANFIS dự báo đa trị tương ứng với giá cổ phiếu. Trên mô hình đề xuất, chương này đã đưa ra hai thuật toán quan trọng đó là thuật toán huấn luyện mạng và thuật toán kiểm thử nhằm minh chứng rõ ràng về cơ sở lý thuyết đã xây dựng, nội dung chương đã thực thi cụ thể một ví dụ thực nghiệm theo từng bước xử lý của mô hình mạng neuro-fuzzy ANFIS đã đề xuất. Theo ví dụ thực nghiệm này cho thấy khả năng hội tụ của thuật toán huấn luyện, kết quả dự báo có độ chính xác cao. Vì vậy, trong chương tiếp theo sẽ thực hiện xây dựng ứng dụng mô phỏng để tạo ra một công cụ dự báo phù hợp và hiệu quả đối với bài toán dự báo giá cổ phiếu.

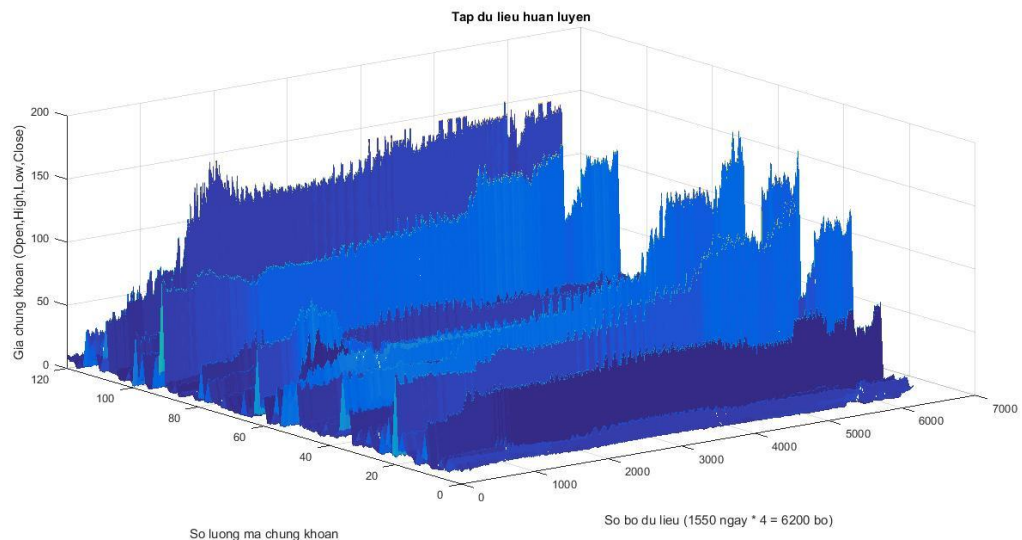
CHƯƠNG 5

XÂY DỰNG ỨNG DỤNG THỰC NGHIỆM

5.1. Giới thiệu

Trong chương này mô tả ứng dụng thực nghiệm của cơ sở lý thuyết đã xây dựng từ Chương 3 và Chương 4. Vì vậy, nội dung của chương thiết kế mô hình thực nghiệm để từ đó làm cơ sở xây dựng ứng dụng mô phỏng về dự báo dữ liệu chuỗi thời gian của thị trường cổ phiếu Cophieu68. Để thực hiện được ứng dụng thực nghiệm, trong chương này thiết kế các hàm chức năng chính dựa trên các thuật toán cơ sở tại Chương 4 dựa trên ngôn ngữ Matlab. Trong thực nghiệm này, luận văn tiến hành thực hiện gồm hai pha: (1) Pha huấn luyện nhằm tìm ra ma trận trọng số suy luận mờ Takagi-Sugeno; (2) Pha kiểm thử thực hiện quá trình dự báo ứng với một bộ kiểm thử đầu vào. Sau khi thực hiện ứng dụng, các số liệu thực nghiệm cũng sẽ được đo đạc trên nền tảng Matlab để từ đó đánh giá hiệu quả của mô hình dự báo đã đề xuất.

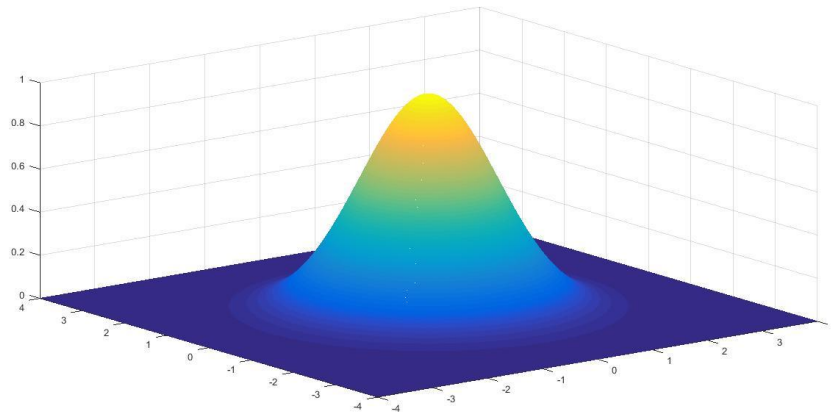
5.2. Mô tả dữ liệu thực nghiệm



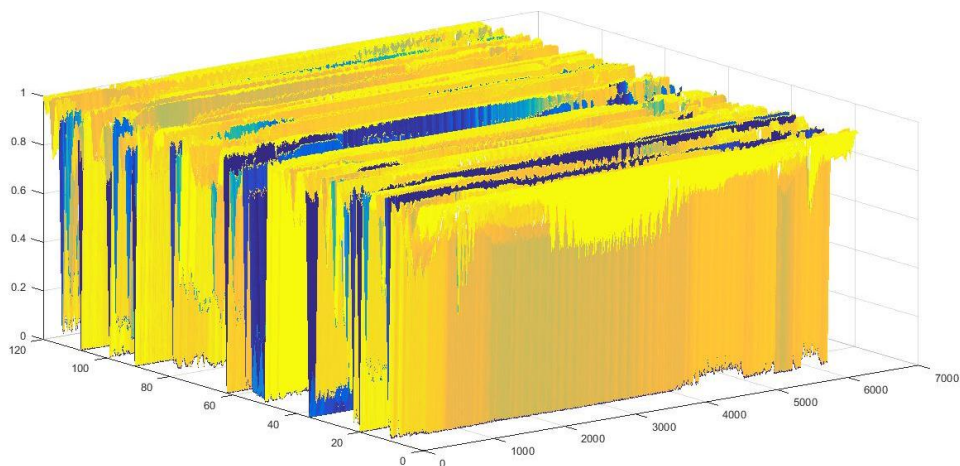
Hình 5.1. Biểu đồ dữ liệu huấn luyện giá cổ phiếu (ĐVT: 1,000VNĐ)

Dữ liệu thực nghiệm được lấy từ cơ sở dữ liệu của Cophieu68 [34] trong khoảng thời gian từ năm 2012 đến năm 2016. Dữ liệu được bố trí thành một ma trận có 6,200 cột ứng với 1,550 ngày liên tiếp gồm 4 trạng thái <Open>, <High>, <Low>, <Close> và 120 dòng ứng với 120 mã cổ phiếu. Trong dữ liệu này trạng thái <Open> nghĩa là

giá khởi điểm của phiên giao dịch trong ngày, trạng thái <High> nghĩa là giá cao nhất của phiên giao dịch trong ngày, trạng thái <Low> nghĩa là giá thấp nhất của phiên giao dịch trong ngày và trạng thái <Close> là giá đóng cửa của phiên giao dịch trong ngày. Dữ liệu được phân đoạn với khung thời gian $\omega = 30$ ngày, mỗi bộ dữ liệu là một ma trận mô tả giá cổ phiếu được giao dịch trong 31 ngày liên tiếp, trong đó ngày cuối cùng là giá cổ phiếu thực tế dùng để huấn luyện hệ suy luận ANFIS. Thực hiện trượt với các khung thời gian khác nhau trên 1,550 ngày liên tiếp để lấy ra 50 bộ dữ liệu huấn luyện. Do đó, ứng với 120 mã cổ phiếu, số lượng mẫu huấn luyện là $31 \times 50 \times 120 = 186,000$ mẫu dữ liệu.

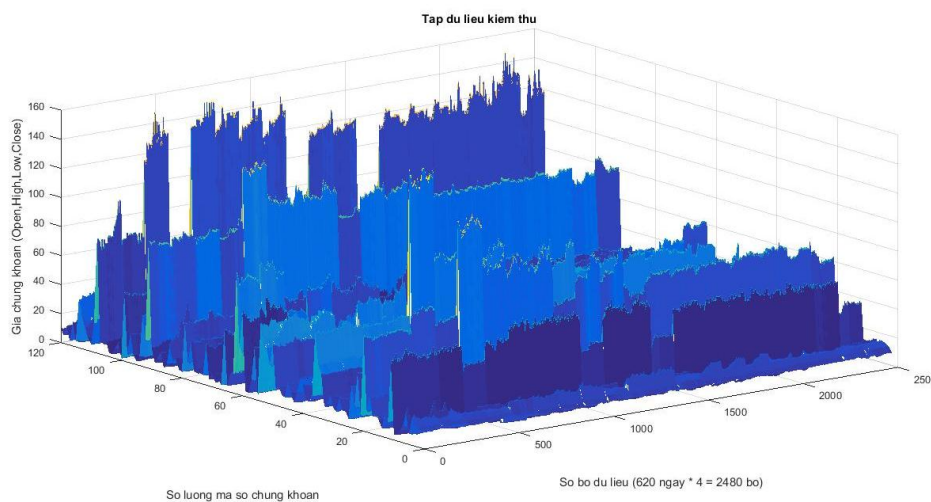


Hình 5.2. Biểu đồ dữ liệu mờ hóa của tập dữ liệu huấn luyện



Hình 5.3. Hàm mờ hóa tập dữ liệu huấn luyện

Bộ dữ liệu kiểm thử cũng được lấy từ cơ sở dữ liệu của Cophieu68 trong khoảng thời gian từ năm 2012 đến năm 2016 bằng cách trượt khung thời gian độc lập với khung thời gian của bộ dữ liệu huấn luyện. Bộ dữ liệu kiểm thử gồm 620 ngày ứng với 120 mã cổ phiếu. Số mẫu kiểm thử là 20 mẫu (chiếm 40% của tổng mẫu dữ liệu). Do đó, số lượng mẫu kiểm thử là $20 \times 31 \times 120 = 74,400$ mẫu dữ liệu. Thực hiện kiểm thử, tính toán sai số dự báo trên bộ dữ liệu này nhằm kiểm tra và đánh giá tính đúng đắn của phương pháp.



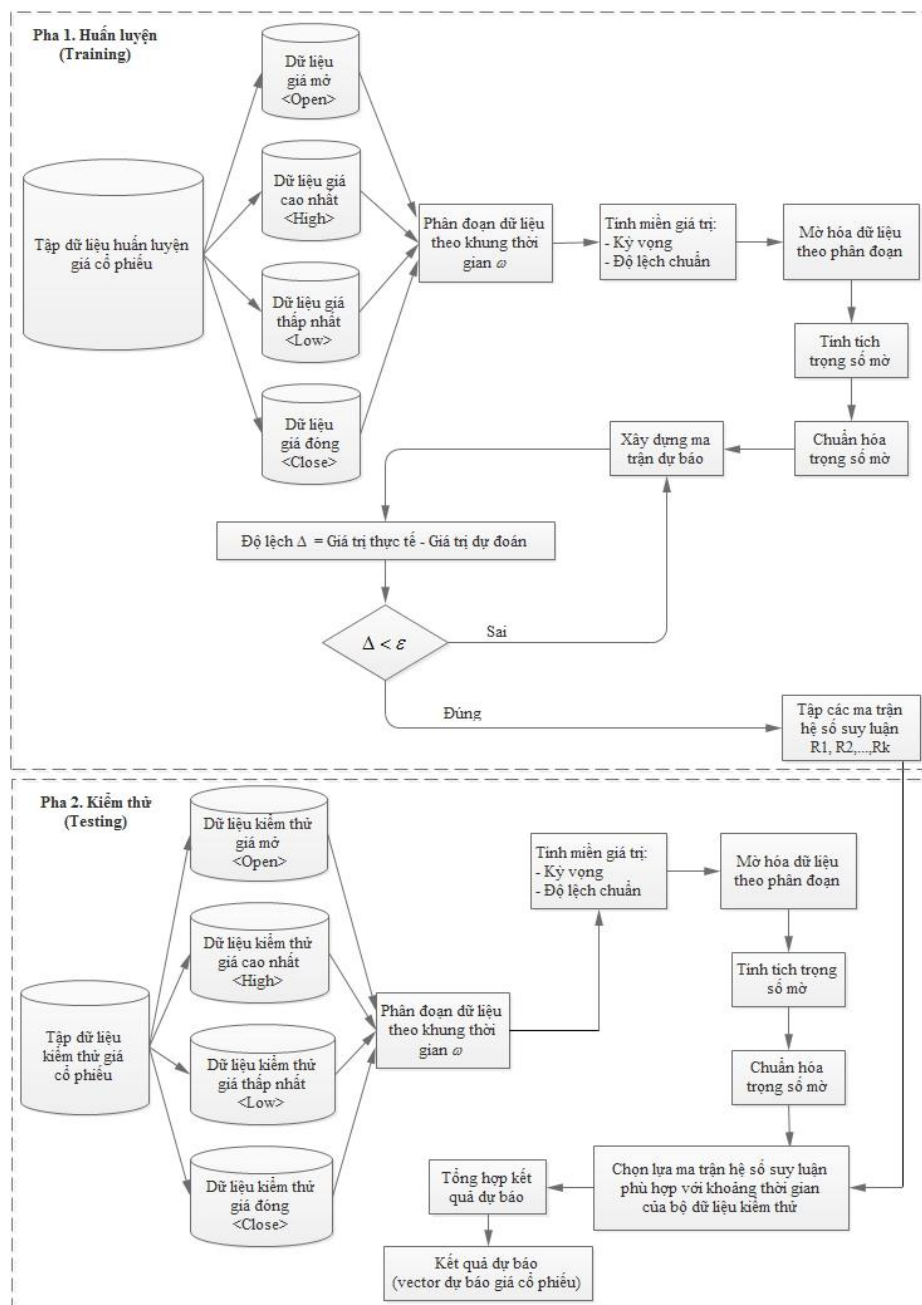
Hình 5.4. Mô tả dữ liệu kiểm thử giá cổ phiếu (ĐVT: 1,000VNĐ)

5.3. Mô tả mô hình thực nghiệm

Trong **Hình 5.5** mô tả mô hình thực nghiệm dự báo giá cổ phiếu. Quá trình dự báo gồm hai pha: (1) Pha huấn luyện dữ liệu cho hệ suy luận ANFIS nhằm tìm ra tập các ma trận hệ số suy luận, chuẩn bị dữ liệu để tạo thành ứng dụng; (2) Pha kiểm thử nhằm kiểm chứng lại quá trình huấn luyện dữ liệu của hệ suy luận ANFIS, sử dụng các dữ liệu đã được chuẩn bị ở pha thứ nhất. Hai pha này thực hiện độc lập với nhau và được mô tả cụ thể như sau:

Pha 1: Pha huấn luyện. Đầu tiên, thực hiện chuẩn hóa tập dữ liệu huấn luyện giá cổ phiếu ứng với khung thời gian ω bất kỳ. Tiến hành mờ hóa tập dữ liệu huấn luyện này bằng phương pháp Gaussian của tập mờ hình chuông dựa trên kỳ vọng μ và phương sai σ trên từng cột dữ liệu. Sau đó, thực hiện tính tích trọng số mờ dựa trên phép tích hợp các số mờ. Chuẩn hóa các số mờ trên đoạn $[0,1]$ nhằm đánh giá mức

độ trọng số của dữ liệu. Xây dựng ma trận dự báo dựa vào kết quả có được từ phép tính tích trọng số mờ và chuẩn hóa. Trên cơ sở ma trận dự báo thực hiện việc huấn luyện dữ liệu. Nếu độ lệch giữa giá trị thực tế và giá trị dự báo lớn hơn sai số ε cho trước thì thực hiện huấn luyện lại cho đến khi giá trị hàm lỗi nhỏ nhất thì dừng lại. Kết quả của quá trình huấn luyện là tập các ma trận hệ số suy luận $R = \{R_1, \dots, R_k\}$ phù hợp với khoảng thời gian của dữ liệu dự báo.



Hình 5.5. Mô hình thực nghiệm dự báo giá cổ phiếu

Pha 2: Pha kiểm thử. Thực hiện chuẩn hóa bộ dữ liệu dữ liệu kiểm thử với khung thời gian ω đã có trong pha huấn luyện. Tiến hành mờ hóa tập dữ liệu kiểm thử này bằng hàm mờ Gaussian hình chuông. Tính tích trọng số mờ dựa trên phép tích hợp các số mờ và chuẩn hóa các số mờ trên đoạn $[0,1]$. Với ma trận hệ số suy luận đã có trong pha huấn luyện thực hiện quá trình huấn luyện dựa trên hệ ANFIS và phương pháp suy luận Takagi-Sugeno sẽ thu được một vector dự báo. Từ đó, có thể dự báo liên tục nhiều ngày tương lai.

5.4. Quá trình thực nghiệm

5.4.1. Các hàm chức năng chính

Dựa trên cơ sở lý thuyết đã đề xuất tại Chương 4 và mô hình thực nghiệm tại Chương 5, luận văn thực hiện thiết kế các hàm chức năng chính dựa trên ngôn ngữ Matlab như sau:

- Dữ liệu huấn luyện được lưu trữ dưới dạng tập tin .csv để dễ dàng ứng dụng trong nhiều dạng thiết bị khác nhau, do đó cần thiết kế một hàm đọc dữ liệu từ file này như sau:

```
Function [{Mi}, Reali] = ReadFileList(filename)
```

- Để thực hiện mờ hóa theo phương pháp mờ Gaussian cần phải tính kỳ vọng và độ lệch chuẩn. Vì vậy, cần thiết kế một hàm tính kỳ vọng và độ lệch chuẩn ứng với từng cột của ma trận đầu vào M_i như sau:

```
Function [Mediani, Stdevi] = MedianStdev(Mi)
```

- Để thực hiện mờ hóa ma trận dữ liệu đầu vào bằng phương pháp Gaussian cần phải tính giá trị hàm Gaussian và ứng dụng để mờ hóa các giá trị ma trận đầu vào M_i . Vì vậy, cần có các hàm như sau:

```
Function fuzzynumber = Gaussian(x, median, stdev)
```

```
Function Mfuzzyi = FuzzyMatrix(Mi)
```

- Mạng neuro-fuzzy ANFIS đã đề xuất cần phải tính trọng số mờ tại tầng 2 và chuẩn hóa trọng số mờ tại tầng 3. Do đó, cần thiết kế hàm tính trọng số mờ cũng như hàm chuẩn hóa trọng số mờ như sau:

```
Function {Mieszkowicz-Rolka, #28} = FuzzyWeight(Mfuzzyi)
```

Function $\{nW_i\} = \text{FuzzyNormalization}(W_i)$

- Để tìm ra được giá trị kết quả dự báo, cần phải tính được ma trận suy luận F_k cũng như chuẩn hóa ma trận này. Vì vậy, cần phải thiết kế hàm tính ma trận suy luận, hàm chuẩn hóa ma trận suy luận và hàm tổng hợp suy luận như sau:

Function $\{F_k\} = \text{ReasoningMatrix}(M_k, R_k)$

Function $\{nF_k\} = \text{NormalizationF}(F_k, nW_k)$

Function $\{f_k\} = \text{Reasoning}(nF_k, R_k)$

- Hàm sau cùng cần phải thiết kế đó là tính sai số bằng trung bình độ lệch. Do đó, hàm tính độ lệch trung bình được thiết kế như sau:

Function $[\text{MeanErr}_k, \text{ErrArr}_k] = \text{DoLech}(f_k, \text{Real}_k);$

5.4.2. Thực nghiệm pha huấn luyện

Trên cơ sở đã thiết kế các hàm như trên và thuật toán huấn luyện đã đề xuất tại Chương 4, trong phần này mô tả việc cài đặt thuật toán huấn luyện như sau:

Đầu vào: Dãy ma trận dữ liệu $\{M_i\}$ mô tả tương ứng các chỉ số cổ phiếu

Đầu ra: Dãy ma trận hệ số suy luận $\{R_i\}$

Bắt đầu

Bước 1: Khởi tạo

- Khởi tạo dãy ma trận $\{M_i\}$:

Lưu file dữ liệu: huanluyen.csv gồm các số liệu cổ phiếu

Đọc dữ liệu từ file huanluyen.csv

Function $[\{M_i\}, \text{Real}_i] = \text{ReadFileList}(\text{filename})$

- Tính kỳ vọng và độ lệch chuẩn từng cột/ mỗi ma trận M_i

Function $[\text{Median}_i, \text{Stdev}_i] = \text{MedianStdev}(M_i)$

- Mờ hóa dãy ma trận $\{M_i\}$ bằng hàm Gaussian:

Function $\text{fuzzynumber} = \text{Gaussian}(x, \text{median}, \text{stdev})$

Function $M_{\text{fuzzy}_i} = \text{FuzzyMatrix}(M_i)$

- Tính trọng số mờ:

Function $\{\text{Mieszkowicz-Rolka}, \#28\} = \text{FuzzyWeight}(M_{\text{fuzzy}_i})$

- Chuẩn hóa tập trọng số mờ:

Function $\{nW_i\} = \text{FuzzyNormalization}(W_i)$

- Phát sinh ngẫu nhiên dãy ma trận hệ số suy luận: R_1, R_2, \dots, R_k

Do đó, phát sinh ngẫu nhiên ma trận $R_{n \times (n+1)}$

Function $\{R_i\} = \text{RandomMatrix}()$

- Khởi tạo sai số $\{ERROR_i\}$;

$ERROR_i, flag_i = 1;$

Bước 2: Huấn luyện

While (Kiemtra(flag_i) = 1) **do**

For $k = 1: K$

If flag_i = 1 **then**

Tính Ma trận suy luận $\{F_k\}$

Function $\{F_k\} = \text{ReasoningMatrix}(M_k, R_k)$

Chuẩn hóa ma trận suy luận F_k

Function $\{nF_k\} = \text{NormalizationF}(F_k, nW_k)$

Tổng hợp suy luận

Function $\{f_k\} = \text{Reasoning}(nF_k, R_k)$

Tính độ lệch

Function $[\text{MeanErr}_k, \text{ErrArr}_k,] = \text{DoLech}(f_k, \text{Real}_k);$

If $\text{MeanErr}_k < \text{ERROR}_k$ **then**

$\text{ERROR}_i = \text{MeanErr}_i;$

Lưu ma trận hệ số R_k ;

Lưu kết quả dự báo và sai số;

$C_k = 0;$

Else

$C_k = C_k + 1;$

End

End

If $C_k > 1000$ **then** flag_k = 0;

End //For

Nếu có thay đổi: Lưu ma trận $\{Aras, \#17\}$

Phát sinh ma trận mới: **Function** $\{Aras, \#17\} = \text{ViPhan}();$

End //While

Return;

Kết thúc

5.4.3. Thực nghiệm pha kiểm thử

Trên cơ sở thuật toán kiểm thử đã đề xuất và dựa trên mô hình mạng neuro-fuzzy ANFIS, trong phần này cài đặt cụ thể thuật toán kiểm thử như sau:

Đầu vào: Dãy dữ liệu kiểm thử $\{T_i\}$, Dãy $\{R_i\}$, Dãy $\{real_i\}$

Đầu ra: Dãy dự báo $\{f_i\}$, Sai số $\{Err_i\}$

Bắt đầu:

Bước 1:

- Đọc dãy ma trận $\{T_i\}$:

Lưu file dữ liệu: `kiemthu.csv` gồm các số liệu cổ phiếu

Đọc dữ liệu từ file `kiemthu.csv`

Function $[T_i, Real_i] = ReadFile(filename)$

- Tính kỳ vọng và độ lệch chuẩn từng cột/ mỗi ma trận M_i

Function $[Median_i, Stdev_i] = MedianStdev(T_i)$

- Mờ hóa dãy ma trận $\{T_i\}$ bằng hàm Gaussian:

Function `fuzzynumber = Gaussian(x, median, stdev)`

Function $Tfuzzy_i = FuzzyMatrix(T_i)$

- Tính trọng số mờ:

Function $\{Mieszkowicz-Rolka, \#28\} = FuzzyWeight(Tfuzzy_i)$

- Chuẩn hóa tập trọng số mờ:

Function $\{nW_i\} = FuzzyNormalization(W_i)$

Bước 2:

For $i = 1: m$

Tính ma trận suy luận $\{F_i\}$

Function $\{F_i\} = ReasoningMatrix(T_i, R_i)$

Chuẩn hóa ma trận suy luận F_i

Function $\{nF_i\} = NormalizationF(F_i, nW_i)$

Tổng hợp suy luận

Function $\{f_i\} = Reasoning(nF_i, R_i)$

Tính độ lệch

Function $[ErrArr_i, MeanErr_i] = DoLech(f_i, Real_i);$

End//For

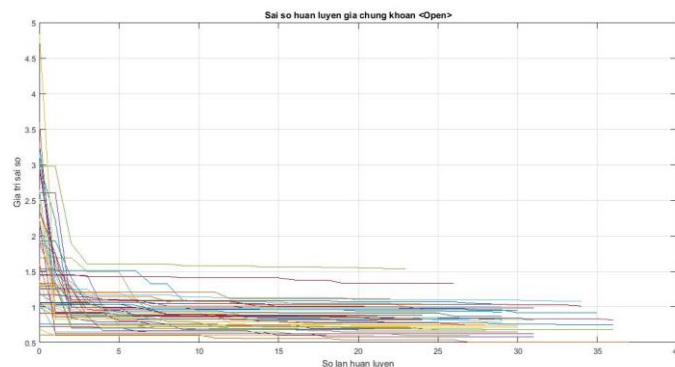
Kết thúc

5.5. Kết quả thực nghiệm và đánh giá kết quả

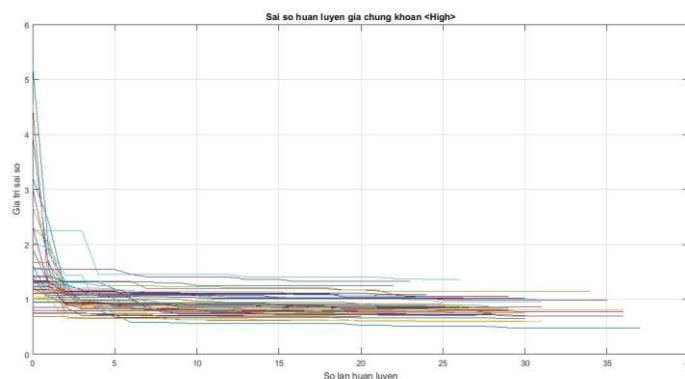
5.5.1. Mô tả kết quả thực nghiệm

Ứng dụng thực nghiệm gồm: pha huấn luyện và pha kiểm thử. Pha huấn luyện dữ liệu được thực thi trên Matlab phiên bản 2013 trên hệ điều hành Windows Server 2008 R2 Enterprise, bộ xử lý Intel(R) Xeon(R) CPU X3440 @ 2.53GHz – Ram 8GB. Pha kiểm thử dữ liệu được thực thi trên Matlab phiên bản 2008 trên hệ điều hành Windows 8.1 Professional, bộ xử lý Core i3 – Ram 4GB. Quá trình thực nghiệm được đo đạc trên đơn vị thời gian là giây theo chuẩn của Matlab. Toàn bộ các hàm hỗ trợ của thuật toán đều được cài đặt cụ thể trên Matlab ngoại trừ các hàm cơ bản. Các số liệu được đo đạc và biểu diễn đồ thị được thực hiện trực tiếp trên Matlab.

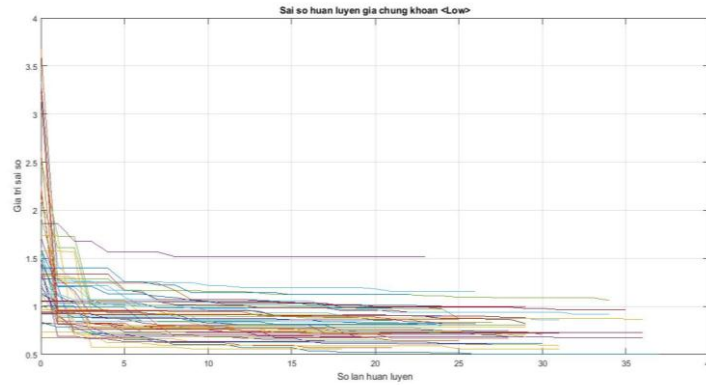
Dựa trên hàm lỗi đã mô tả như trên, luận văn đánh giá sai số trung bình huấn luyện được mô tả như sau:



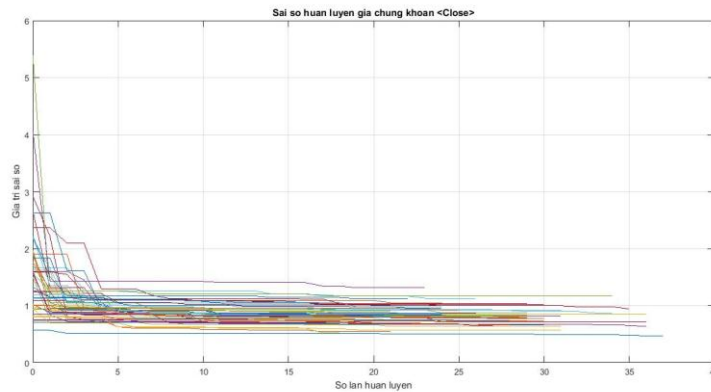
Hình 5.6. Sai số huấn luyện giá cổ phiếu với trạng thái <Open>



Hình 5.7. Sai số huấn luyện giá cổ phiếu với trạng thái <High>



Hình 5.8. Sai số huấn luyện giá cổ phiếu với trạng thái <Low>

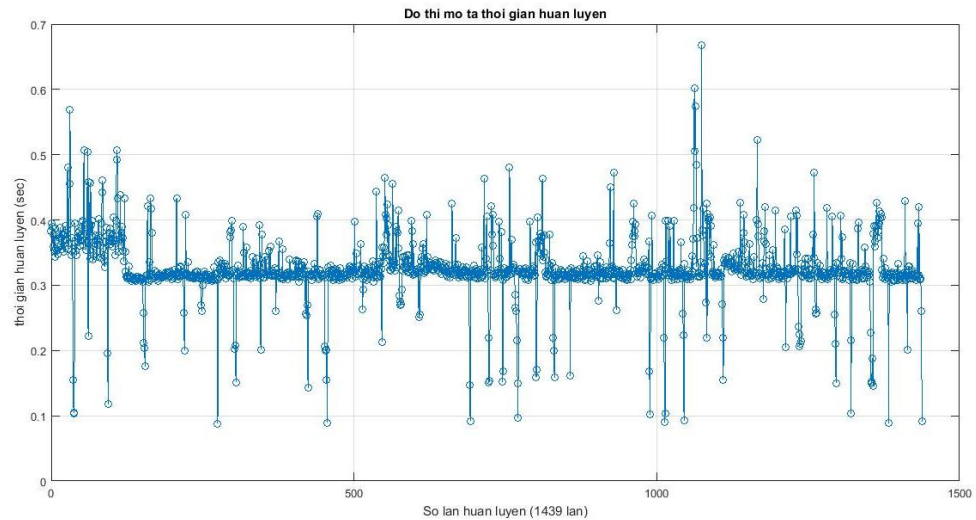


Hình 5.9. Sai số huấn luyện giá cổ phiếu với trạng thái <Close>

Từ **Hình 5.6** đến **Hình 5.9** mô tả sai số huấn luyện hệ ANFIS ứng với dữ liệu Cophieu68 tại các trạng thái giá cổ phiếu gồm <Open>, <High>, <Low>, <Close>. Theo đồ thị cho thấy tốc độ hội tụ của hệ ANFIS tương đối nhanh. Từ đó cho thấy mô hình đã đề xuất phù hợp với bài toán dự báo dữ liệu dạng chuỗi thời gian. Số lần huấn luyện nhiều nhất là 39 lần, số lần huấn luyện thấp nhất là 19 lần tương ứng với các bộ dữ liệu huấn luyện M_1 và M_{11} , số lần huấn luyện trung bình xấp xỉ 29 lần tương ứng với 50 bộ dữ liệu huấn luyện từ M_1 đến M_{50} .

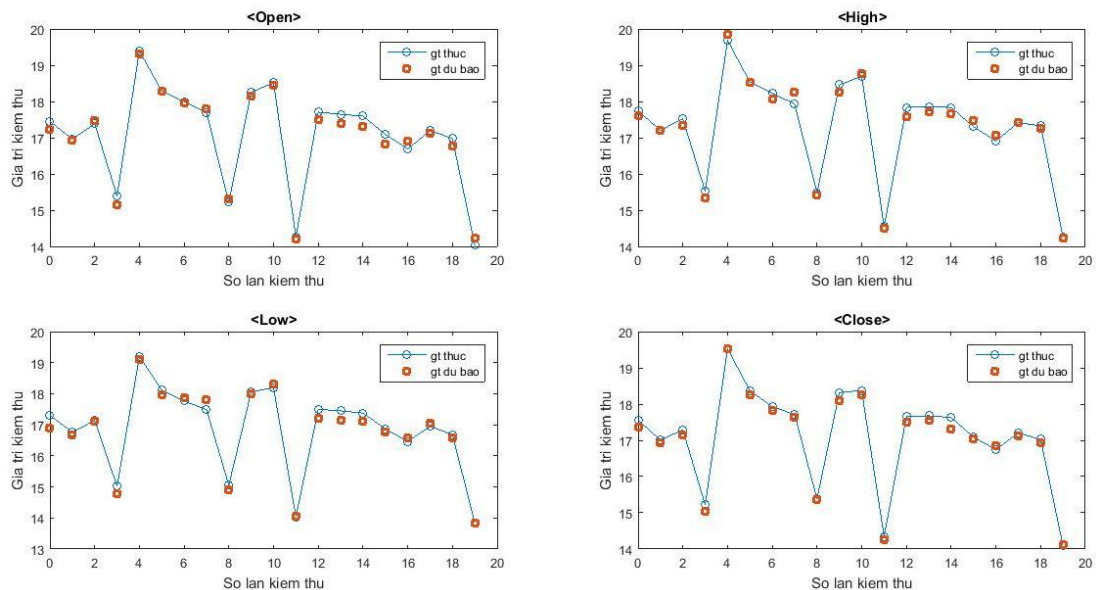
Bảng 5.1. Đánh giá sai số huấn luyện bộ dữ liệu từ M_1 đến M_{50}

Trạng thái giá cổ phiếu	Sai số hội tụ nhỏ nhất	Sai số hội tụ lớn nhất	Sai số hội tụ trung bình
Open	0.50265	1.5362	0.8374516
High	0.48421	1.3555	0.8479776
Low	0.50173	1.5174	0.8061498
Close	0.46963	1.3114	0.8232176



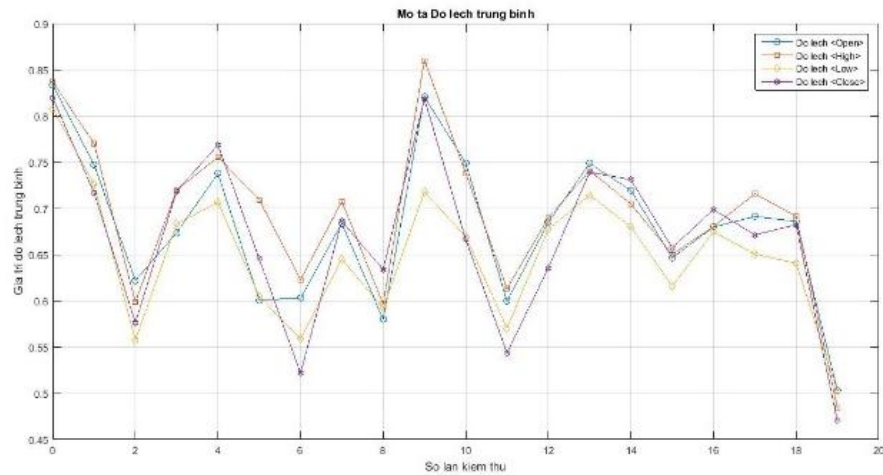
Hình 5.10. Thời gian huấn luyện của toàn bộ tập dữ liệu

Hình 5.10 mô tả toàn bộ thời gian huấn luyện của quá trình huấn luyện mạng neuro-fuzzy ANFIS. Tổng số lần huấn luyện tất cả các mẫu dữ liệu là 1,439 lần. Tổng thời gian huấn luyện là 467.2285 giây, tức là 7.787142 phút. Trong đó, thời gian huấn luyện cao nhất là 0.6671 giây, thời gian huấn luyện thấp nhất là 0.08821 giây, thời gian huấn luyện trung bình là 0.32469 giây. Qua số liệu thống kê về thời gian huấn luyện cho thấy rằng mỗi lần huấn luyện có tốc độ tương đối nhanh, do đó việc huấn luyện theo thuật toán đã đề xuất có tính khả thi cao.



Hình 5.11. Mô tả giá trị dự báo và giá trị thực

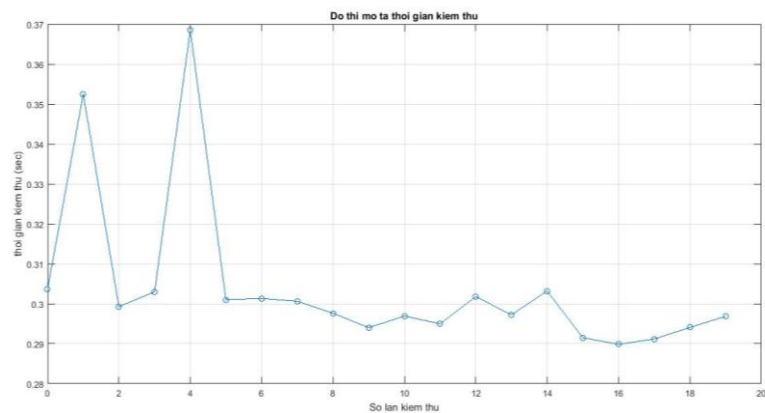
Sau khi thực hiện huấn luyện hệ ANFIS, luận văn tiến hành quá trình kiểm thử nhằm dự báo kết quả. Việc kiểm thử này được thực hiện trên bài toán dự báo giá cổ phiếu nhằm minh chứng tính tổng quát và tính đúng đắn của mô hình đã đề xuất. Tại **Hình 5.11** mô tả giá trị thực và giá trị dự báo của giá cổ phiếu ứng với bốn trạng thái: Giá mở <Open>, giá cao nhất <High>, giá thấp nhất <Low>, giá đóng <Close>.



Hình 5.12. Độ lệch trung bình giữa giá trị dự báo và giá trị thực tế

Bảng 5.2. Đánh giá độ lệch giữa giá trị dự báo và giá trị thực tế

Trạng thái giá cổ phiếu	Sai số dự báo nhỏ nhất	Sai số dự báo lớn nhất	Sai số dự báo trung bình
Open	0.50265368	0.833208491	0.680338622
High	0.484214596	0.860077288	0.694128424
Low	0.501733124	0.807624221	0.649539821
Close	0.469622451	0.819465182	0.669867933



Hình 5.13. Đồ thị mô tả thời gian dự báo

Thời gian dự báo được đo đạc trực tiếp bằng phần mềm Matlab nhằm đánh giá chi phí về tốc độ dự báo nhằm phục vụ cho quá trình định hướng kinh doanh. Đồ thị mô tả thời gian dự báo tại **Hình 5.13**. Trong đồ thị này, mô tả thời gian dự báo ứng với 20 bộ kiểm thử, mỗi lần kiểm thử được đo đạc trên phần mềm Matlab theo đơn vị tính thời gian là giây. Theo kết quả đồ thị này cho thấy thời gian dự báo cao nhất xấp xỉ khoảng 0.37 giây, thời gian dự báo thấp nhất xấp xỉ khoảng 0.29 giây. Hầu hết các khoảng thời gian dự báo biến thiên trong khoảng thời gian từ 0.29 giây đến 0.31 giây, vì vậy tốc độ dự báo của phần mềm dựa trên phương pháp đề xuất là khả thi và đáp ứng được cho việc dự báo nhanh kết quả giá cổ phiếu.

5.5.2. So sánh kết quả

Theo tài liệu [33] vào năm 2015 của F. Zhang, tài liệu [19] vào năm 2016 của Thangjam Ravichandra, tài liệu [18] vào năm 2013 của T. Nguyen và tài liệu [5] vào năm 2015 của Mu-Yen Chen đã thực hiện đánh giá việc dự báo giá cổ phiếu bằng nhiều phương pháp khác nhau. Theo thực nghiệm, luận văn sẽ đánh giá sai số dự báo của phương pháp đề xuất và sai số của các phương pháp khác. Các sai số này bao gồm: MSE, MAE, MAPE, tỉ lệ % MAE, RMSE, MPE.

$$MAE = \frac{1}{n} \sum_{s=1}^n \left| y_s - \hat{y}_s \right| \quad (5.6)$$

$$MAPE = \frac{1}{n} \sum_{s=1}^n \left| \frac{y_s - \hat{y}_s}{y_s} \right| \quad (5.7)$$

$$MSE = \frac{1}{n} \sum_{s=1}^n (y_s - \hat{y}_s)^2 \quad (5.8)$$

Với n là số lượng mã cổ phiếu, y_s là giá trị thực tế của cổ phiếu, \hat{y}_s là giá trị dự báo của cổ phiếu.

Bảng 5.3. So sánh sai số giữa phương pháp đề xuất và các phương pháp khác

Loại sai số Phương pháp	MSE	MAE	MAPE
T-Norm dựa trên ANFIS [33]	1.8069	1.3372	0.2534
ANFIS truyền thống [33]	2.3729	1.7263	0.4012
ANFIS đề xuất	0.0296; 0.0233; 0.3079; 0.1087	0.1474; 0.1282; 0.5268; 0.3087	0.0088; 0.0073; 0.0306; 0.0178

Trong **Bảng 5.3** đánh giá các sai số gồm MSE, MAE và MAPE giữa phương pháp đề xuất và phương pháp dựa trên ANFIS T-Norm và ANFIS truyền thống. Đối với sai số MAE thì phương pháp đề xuất nhỏ hơn hai phương pháp trên. Điều này cho thấy sự chính xác của pha kiểm thử. Tương tự, đối với các sai số MSE và MAPE thì phương pháp đề xuất cũng có sai số thấp hơn.

$$\%MAE = \frac{MAE}{r} \quad (5.9)$$

Với MAE là độ lệch trung bình, r là giá trị thực tế của cổ phiếu.

Bảng 5.4. So sánh tỉ lệ sai số giữa phương pháp đề xuất và phương pháp ANN

Loại sai số Phương pháp	MAE (%)
ANN [19]	7.0
ANFIS đề xuất	3.98; 4.01; 3.86; 3.92

Luận văn cũng thực hiện đánh giá sai số theo tỉ lệ % giữa phương pháp đề xuất và phương pháp ANN. Theo như bảng so sánh trên cho thấy phương pháp đề xuất có tỉ lệ sai số thấp hơn (xấp xỉ thấp hơn khoảng 1.75 lần).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2} \quad (5.10)$$

Với N là số lượng mã cổ phiếu, y_i là giá trị thực tế của cổ phiếu, \tilde{y}_i là giá trị dự báo của cổ phiếu.

Bảng 5.5. So sánh sai số RMSE giữa phương pháp đề xuất và phương pháp mạng nơron mờ loại 2

Loại sai số Phương pháp	RMSE	Sai số RMSE trung bình
NN Fuzzy [18]	0.3854; 0.3347; 0.8367; 0.7298; 0.3656; 0.3565; 0.4582; 0.4494; 0.345; 0.3155; 0.5317; 0.3237	0.452683333
ANFIS đề xuất	0.172167; 0.152791; 0.190428; 0.135342	0.162682122

Ngoài ra, để minh chứng tính hiệu quả của phương pháp đề xuất, trong phần này cũng so sánh các sai số RMSE giữa các trạng thái giá cổ phiếu. Theo như bảng

trên cho thấy sai số RMSE trung bình thực nghiệm của phương pháp đề xuất nhỏ hơn (xấp xỉ 2.8 lần).

Bảng 5.6. So sánh sai số giữa phương pháp đề xuất và các phương pháp khác

Loại sai số Phương pháp	RMSE	MSE	MAE	MPE
Fuzzy GJR-GARCH [5]	0.4453; 0.7809	0.1983; 0.6099	0.3652; 0.7912	0.3795; 0.2852
GARCH [5]	0.7552; 1.2036	0.5704; 1.4487	0.7076; 1.2403	0.8366; 0.6723
GJR-GARCH [5]	0.7641; 1.1928	0.5839; 1.423	0.7298; 1.1955	0.842; 0.6511
Hybird fuzzy (1) [5]	0.3377; 0.5833	0.114; 0.3402	0.2197; 0.6082	0.2614; 0.1591
Hybird fuzzy (2) [5]	0.307; 0.5821	0.0942; 0.3388	0.1676; 0.6058	0.1986; 0.1583
ANFIS đề xuất	0.172167; 0.152791; 0.190428; 0.135342	0.0296; 0.0233; 0.3079; 0.1087	0.1474; 0.1282; 0.5268; 0.3087	0.008758; 0.007329; 0.008986; 0.00665

Bên cạnh đó, luận văn cũng so sánh với nhiều phương pháp khác nhau trong bài toán dự báo giá cổ phiếu. Trong các phương pháp này có thể tiếp cận dưới dạng mạng neuro-fuzzy hoặc bằng một cách tiếp cận khác nhưng đều giải quyết bài toán dự báo giá cổ phiếu. Theo số liệu sai số cho thấy phương pháp dùng mạng neuro-fuzzy ANFIS đã đề xuất đã cải tiến sai số đáng kể so với các phương pháp khác.

5.6. Kết luận

Trong chương này đã tiếp cận xây dựng mô hình ứng dụng thực nghiệm dựa trên mô hình neuro-fuzzy ANFIS đã đề xuất trong Chương 4 và các thuật toán huấn luyện, kiểm thử tương ứng. Trên cơ sở này, nội dung của chương đã trình bày quá trình cài đặt cụ thể của các thuật toán huấn luyện và kiểm thử trên ngôn ngữ Matlab. Theo ứng dụng thực nghiệm cho thấy phương pháp đề xuất đã cho các kết quả dự báo chính xác và hiệu quả, tức là thời gian huấn luyện tương đối thấp, thời gian kiểm thử ngắn đồng thời có độ chính xác cao. Luận văn đã giải quyết được vấn đề dự báo đồng thời nhiều kết quả, giúp cải thiện đáng kể về tốc độ so với các phương pháp khác. Điều này đã được minh chứng trong phần kết quả thực nghiệm. Hướng phát triển của luận văn là đề xuất hệ ANFIS có khả năng tự nhận biết các trường hợp dị biệt và có khả năng chịu lỗi cao để kết quả dự báo không bị ảnh hưởng bởi các giá trị bất thường.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong luận văn đã tiếp cận giải quyết bài toán dự báo giá cổ phiếu dựa trên mô hình mạng ANFIS đã đề xuất. Đồng thời, luận văn cũng đã xây dựng ứng dụng thực nghiệm nhằm minh chứng cho cơ sở lý thuyết đã đề nghị và thực hiện đánh giá giữa phương pháp đề xuất và các phương pháp khác cùng thể loại cũng như không cùng thể loại. Trong luận văn đã tiếp cận hệ suy luận mờ Takagi-Sugeno và mạng nơron nhân tạo ANN làm cơ sở lý thuyết để từ đó tiếp cận mô hình mạng neuro-fuzzy ANFIS. Dựa trên bộ dữ liệu Cophieu68, trong luận văn đã đề xuất một mô hình dự báo tổng quát đa trị để từ đó dự báo giá cổ phiếu đồng thời tương ứng với từng loại mã cổ phiếu. Trên cơ sở mô hình đã đề xuất, luận văn đã trình bày chi tiết thuật toán huấn luyện cũng như thuật toán kiểm thử, từ đó thực thi một ví dụ thực nghiệm trên bộ dữ liệu thực tế. Dựa trên công cụ lập trình Matlab, luận văn đã trình bày chi tiết về quá trình cài đặt mô hình dự báo cũng như thuật toán huấn luyện và thuật toán kiểm thử cho mạng neuro-fuzzy ANFIS. Trên cơ sở lý thuyết đã đề nghị, luận văn đã phân tích bộ dữ liệu Cophieu68 để từ đó áp dụng vào mô hình cũng như các thuật toán đã đề xuất. Trong Chương 5 đã xây dựng ứng dụng thực nghiệm và đánh giá kết quả của mô hình đề xuất trên bộ dữ liệu thực tế. Theo thực nghiệm cho thấy kết quả huấn luyện có thời gian huấn luyện và tốc độ hội tụ tương đối nhanh, điều này cho thấy tính hiệu quả của quá trình huấn luyện mạng neuro-fuzzy ANFIS đã đề xuất. Ngoài ra, quá trình kiểm thử trên các bộ dữ liệu độc lập có thời gian tương đối thấp, tức là quá trình dự báo thực hiện hiệu quả. Luận văn tiến hành so sánh phương pháp đề xuất với các phương pháp đồng dạng cũng như các phương pháp không đồng dạng nhưng đều giải quyết bài toán giá cổ phiếu. Theo kết quả so sánh, phương pháp đề xuất hầu hết có sai số dự báo thấp hơn. Điều này cho thấy phương pháp đề xuất trong luận văn thể hiện tính hiệu quả cũng như tính chính xác đối với bài toán dự báo giá cổ phiếu. Trong phần cải tiến tiếp theo của luận văn sẽ xây dựng mô hình dự báo dữ liệu chuỗi thời gian của nhiều ngày tương lai. Ngoài ra, sẽ phát triển mạng neuro-fuzzy ANFIS tự nhận biết các trường hợp dị biệt và có khả năng chịu lỗi cao để kết quả dự báo không bị ảnh hưởng bởi các giá trị bất thường này.

DANH MỤC CÔNG TRÌNH LIÊN QUAN CỦA TÁC GIẢ ĐÃ CÔNG BỐ

- [1] *Hệ ANFIS cho bài toán dự báo dữ liệu chuỗi thời gian và ứng dụng*, Hội nghị Khoa học Quốc gia lần thứ 9, Trường Đại học Cần Thơ, Kỷ yếu Hội thảo Quốc gia về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin, 2016. [**Accepted**]
- [2] *FS-Alg: Thuật toán khai thác dãy sự kiện phổ biến*, Hội thảo toàn quốc về CNTT & Truyền thông 2015, Trường Đại học Cần Thơ, 07/11/2015, Tạp chí Khoa học Trường Đại học Cần Thơ, Số chuyên đề: Công nghệ Thông tin (2015), Tr. 128-135, 2015.
- [3] *Khai thác dãy sự kiện phổ biến sử dụng cây Seq-Tree*, Hội thảo Khoa học SAICT'2015, Trường Đại học Khoa học – Đại học Huế, 30/05/2015, Tạp chí Khoa học Đại học Huế, Chuyên san Kỹ thuật và Công nghệ, ISSN 1859-1388, Tập 106, Số 07, Tr. 109-120, 2015.

TÀI LIỆU THAM KHẢO

- [1] A. Al-Hmouz, et al. (2012). 'Modeling and Simulation of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for Mobile Learning'. *IEEE transactions on learning technologies*. 5 (3), 226-237.
- [2] R. Lee (2016). *Detection of Dengue Epidemic in Dhaka, Bangladesh by a Neuro Fuzzy Approach*, M. Arifuzzaman, et al., In: *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. Springer Switzerland. pp. 165-174.
- [3] S. Campisi-Pinto, J. Adamowski, G. Oron (2013). 'Forecasting Urban Water Demand Via Wavelet-Denoising and Neural Network Models. Case Study: City of Syracuse, Italy'. *Water Resources Management*. 26 (12), 3539-3558.
- [4] C.L.P. Chen, Chi-Hsu Wang (2014). 'A New Learning Algorithm for a Fully Connected Neuro-Fuzzy Inference System'. *IEEE transactions on Neural networks and learning systems*. 25 (10), 1741-1757.
- [5] M.-Y. Chen, B.-T. Chen (2015). 'A hybrid fuzzy time series model based on granular computing for stock price forecasting'. *Information Sciences*. 294, 227–241.
- [6] W. Pedrycz, S.-M. Chen (2013). *A Best-Match Forecasting Model for High-Order Fuzzy Time Series*, Y.-C. Cheng, S.-T. Li, In: *Time Series Analysis, Modeling and Applications*. Springer Berlin pp. 331-345.
- [7] P. Esling, C. Agon (2012). 'Time-Series Data Mining'. *ACM Computing Surveys*. 45 (1).
- [8] T.-c. Fu (2011). 'A review on time series data mining'. *Engineering Applications of Artificial Intelligence*. 24 (1), 164–181.
- [9] C. Mao, T. Yokomori (2006). *Fuzzy Forecasting with DNA Computing*, D.J.-F. Jeng, et al., In: *DNA Computing*. Springer Berlin. pp. 324-336.
- [10] L. Jing (2013). 'A neural fuzzy inference system'. *Journal of electronics*. 30 (4), 401-410.

- [11] K. Deep, *et al.* (2012). *A Computational Method of Forecasting Based on Intuitionistic Fuzzy Sets and Fuzzy Time Series*, B.P. Joshi, S. Kumar, In: *Advances in Intelligent and Soft Computing*. Springer India. pp. 993-1000.
- [12] S. Kara, S. Dasb, P.K. Ghosh (2014). 'Applications of neuro fuzzy systems: A brief review and future outline'. *Applied Soft Computing*. 15, 243–259.
- [13] Y. Liao, *et al.* (2013). 'Stock Price Forecast Using Tree Augmented Naïve (TAN) Bayes'. *Advances in Intelligent Systems and Computing*. 212, 1013-1019.
- [14] O. Maimon, L. Rokach. (2005). *Data mining and knowledge discovery handbook*. Springer. New York.
- [15] K.N. Das, *et al.* (2015). *Vehicle Classification Using Adaptive Neuro-Fuzzy Inference System (ANFIS)*, A.K. Maurya, D.K. Patel, In: *Proceedings of Fourth International Conference on Soft Computing for Problem Solving Book Subt.* Springer India. pp. 137-152.
- [16] L. RUTKOWSKI, *et al.* (2009). *Neuro-fuzzy Systems*, R.A. Meyers, In: *Computational Complexity*. Springer: New York. pp. 2069-2081.
- [17] C.G. Helmis, P.T. Nastos (2013). *24-Hours Ahead Forecasting of PM10 Concentrations Using Artificial Neural Networks in the Greater Athens Area, Greece*, K. Moustris, *et al.*, In: *Advances in Meteorology, Climatology and Atmospheric Physics*. Springer Berlin pp. 1121-1126.
- [18] T. Nguyen, *et al.* (2013). *Neural Network and Interval Type-2 Fuzzy System for Stock Price Forecasting* Paper presented at IEEE International Conference on Fuzzy Systems (FUZZ), Hyderabad.
- [19] T. Ravichandra, C. Thingom (2016). 'Stock Price Forecasting Using ANN Method'. *Advances in Intelligent Systems and Computing*. 435, 599-605.
- [20] I.E. Shepelev, *et al.* (2015). 'Comparing Ambient Temperature Account Methods in Neural Network Based City Short -Term Load Forecasting'. *Optical Memory and Neural Networks*. 24 (3), 220–229.

- [21] P. Singh (2016). *Two-Factors High-Order Neuro-Fuzzy Forecasting Model*, P. Singh, In: *Applications of Soft Computing in Time Series Forecasting*. Springer Switzerland. pp. 83-97.
- [22] V. Vijay, et al. (2015). *Ground-Based Measurement for Solar Power Variability Forecasting Modeling Using Generalized Neural Network*, V.P. Singh, et al., In: *Systems Thinking Approach for Social Problems*. Springer India. pp. 49-61.
- [23] I. Svalina, et al. (2013). 'An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices'. *Expert Systems with Applications*. 40 (15), 6055–6063.
- [24] V.-H. Nguyen, A.-C. Le, V.-N. Huynh (2015). *A New Approach to Multi-variable Fuzzy Forecasting Using Picture Fuzzy Clustering and Picture Fuzzy Rule Interpolation Method*, P.H. Thong, L.H. Son, In: *Knowledge and Systems Engineering*. Springer: New York. pp. 679-690.
- [25] B.V. Babu, et al. (2014). *Fault Diagnosis of Ball Bearings Using Support Vector Machine and Adaptive Neuro Fuzzy Classifier*, R. Tiwari, P.K. Kankar, V.K. Gupta, In: *Advances in Intelligent Systems and Computing*. Springer India. pp. 1477-1482.
- [26] K. Vasimalla (2014). 'A Survey on Time Series Data Mining '. *International Journal of Innovative Research in Computer and Communication Engineering*. 2 (5), 170-179.
- [27] V. Golovko, A. Imada. (2014). *Neural Networks and Artificial Intelligence*. Springer. New York Dordrecht London.
- [28] V.N. Huynh, T.B.Ho, Y.Nakamori (2002), *A parametric representation of linguistic hedges in Zadeh's fuzzy logic*, *International Journal of Approximate Reasoning*.
- [29] H. Cartwright. (2015). *Artificial Neural Networks*. New York.
- [30] S.Hoelldobler, N. H. Nga, T. D. Khang (2005), *The fuzzy logic description logic ALC_FLH*, Proceedings of International Workshop on Description Logics – DL2005, Edinburgh, Scotland.

- [31] P. Koprinkova-Hristova, V. Mladenov, N.K. Kasabov. (2015). *Artificial Neural Networks*. Springer. New York.
- [32] Nguyen Cat Ho and Wechler, Hedge algebra (1990), *An algebraic approach to structure of set of linguistic truth values*, Fuzzy Set and system.
- [33] F. Zhang, Z. Liao (2015). *Stock Price Forecasting Based on Multi-Input Hamacher T-Norm and ANFIS*. Paper presented at The Ninth International Conference on Management Science and Engineering Management, July 21 to 23, 2015, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.
- [34] <http://www.cophieu68.vn/>. 2016.