

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN THỊ THÚY LAN

**CÁC PHƯƠNG PHÁP CHIA KHOẢNG TRONG MÔ HÌNH
CHUỖI THỜI GIAN MỜ**

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH

MÃ SỐ: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. NGUYỄN CÔNG ĐIỀU

THÁI NGUYÊN - 2012

LỜI CAM ĐOAN

Tôi 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.

Tác giả luận văn

Nguyễn Thị Thúy Lan

MỤC LỤC

Trang

Trang phụ bìa	
Lời cam đoan	
MỤC LỤC.....	i
MỞ ĐẦU.....	1
CHƯƠNG 1: CÁC KIẾN THỨC CƠ BẢN VỀ CHUỖI THỜI GIAN	4
VÀ TẬP MỜ.....	4
1.1. Chuỗi thời gian và quá trình ngẫu nhiên.....	4
1.1.1. Khái niệm chuỗi thời gian và quá trình ngẫu nhiên	4
1.1.2. Quá trình ngẫu nhiên dừng.....	5
1.1.3. Hàm tự tương quan	6
1.1.4. Toán tử tiến, toán tử lùi.....	7
1.2. Mô hình ARMA.....	7
1.2.1. Quá trình tự hồi quy.....	7
1.2.2. Quá trình trung bình trượt.....	9
1.2.3. Quá trình tự hồi quy trung bình trượt.....	11
1.3. Những hạn chế của mô hình ARMA trong chuỗi thời gian tài chính.....	13
1.4. Lý thuyết tập mờ	16
1.4.1. Tập mờ.....	16
1.4.2. Các phép toán trên tập mờ.....	18
1.5. Các quan hệ và suy luận xấp xỉ, suy diễn mờ.....	21
1.5.1. Quan hệ mờ.....	21
1.5.2. Suy luận xấp xỉ và suy diễn mờ.....	22
1.6. Hệ mờ.....	24
1.6.1. Bộ mờ hoá.....	24
1.6.2. Hệ luật mờ.....	25
1.6.3. Động cơ suy diễn.....	25
1.6.4. Bộ giải mờ.....	26
CHƯƠNG 2: MÔ HÌNH CHUỖI THỜI GIAN MỜ	28
2.1. Chuỗi thời gian mờ	28

2.1.1. Một số khái niệm cơ bản	28
2.1.2. Một số định nghĩa liên quan đến chuỗi thời gian mờ	29
2.2. Một số thuật toán trong mô hình chuỗi thời gian mờ	30
2.2.1. Thuật toán của Song & Chissom [5]	30
2.2.2. Thuật toán của Chen [6]	31
2.2.3. Thuật toán Heuristic của Huarng [9]	31
CHƯƠNG 3: CÁC PHƯƠNG PHÁP CHIA KHOẢNG TRONG MÔ HÌNH	
CHUỖI THỜI GIAN MỜ	33
3.1. Phương pháp chia khoảng	33
3.1.1. Phương pháp lựa chọn ngẫu nhiên	34
3.1.2. Phương pháp độ dài dựa trên sự phân bố giá trị (Huarng [9])	34
3.1.3. Phương pháp độ dài dựa trên giá trị trung bình (Huarng [9])	35
3.1.4. Phương pháp dựa trên mật độ [2]	35
3.2. Ứng dụng trong dự báo.	37
3.2.1. Dự báo chỉ số chứng khoán Đài Loan TAIFEX [8,9].	37
3.2.2. Dự báo chỉ số VN-Index ở Việt Nam	52
KẾT LUẬN	67
TÀI LIỆU THAM KHẢO	68

DANH MỤC HÌNH VẼ, BẢNG BIỂU

Hình 1.1. Chuỗi giá	13
Hình 1.2. Chuỗi tăng trưởng	14
Hình 1.3. Nhiều	14
Hình 1.4. Tự tương quan của nhiều.....	15
Hình 1.5. Tự tương quan riêng của nhiều	15
Hình 1.6. Bình phương nhiều.....	15
Hình 1.7. Tự tương quan bình phương nhiều.....	16
Hình 1.8. Tự tương quan riêng bình phương nhiều	16
Hình 1.9. Hàm liên thuộc của tập mờ “x gần 1”	17
Hình 1.10. Một số dạng hàm liên thuộc của tập mờ	18
Bảng 1.1. Một số phép kéo theo mờ thông dụng	20
Hình 1.11. Cấu hình cơ bản của hệ mờ.....	24
Bảng 3.1. Cơ sở ánh xạ	35
Bảng 3.2. Giá trị chỉ số chứng khoán Đài Loan.....	38
Bảng 3.3. Nhóm mối quan hệ mờ	39
Bảng 3.4. Giá trị mờ và kết quả dự báo	40
Bảng 3.5. Tính giá trị tuyệt đối của hiệu số bậc 1.....	41
Bảng 3.6. Sự phân phối tích lũy của sai phân cấp một	42
Bảng 3.7. Nhóm mối quan hệ mờ	44
Bảng 3.8. Kết quả dự báo.....	45
Bảng 3.9. Nhóm mối quan hệ mờ	47
Bảng 3.10. Kết quả dự báo.....	47
Bảng 3.11. So sánh với các phương pháp dự báo khác.....	49
Hình 3.1. Đồ thị so sánh các kết quả dự báo chỉ số chứng khoán với giá trị thực....	51
Bảng 3.12. Số liệu chỉ số VN-index trong tháng 4 và tháng 5 năm 2012.....	52
Bảng 3.13. Phân bố giá trị trong từng khoảng	53
Bảng 3.14. Phân khoảng.....	54
Bảng 3.15. Nhóm mối quan hệ mờ	55

Bảng 3.16. Nhóm quan hệ mờ và nhóm quan hệ mờ heuristic và dự báo	56
Hình 3.2. Đồ thị so sánh kết quả dự báo bằng phương pháp dựa trên mật độ	58
và giá trị thực.....	58
Bảng 3.17. Tính giá trị tuyệt đối của hiệu số bậc 1.....	58
Bảng 3.18. Sự phân phối tích lũy của sai phân cấp một	59
Bảng 3.19. Nhóm mối quan hệ mờ	61
Bảng 3.20. Kết quả dự báo.....	61
Bảng 3.21. Nhóm mối quan hệ mờ	63
Bảng 3.22. Kết quả dự báo.....	64
Bảng 3.23. So sánh hiệu quả của các thuật toán	65
Hình 3.3. Đồ thị so sánh các kết quả dự báo chỉ số VN-index với giá trị thực	65

MỞ ĐẦU

Chuỗi thời gian mờ và mô hình chuỗi thời gian mờ bậc nhất do Song và Chissom [1] phát triển từ năm 1993. Sau công trình này, một loạt các bài báo của nhiều tác giả khác nhau tiếp tục dựa trên ý tưởng này để dự báo chuỗi thời gian và ứng dụng trong nhiều lĩnh vực khác nhau như dự báo dân số, tài chính, nhiệt độ, nhu cầu điện, vv... Gần đây có rất nhiều tác giả liên tục cải tiến mô hình chuỗi thời gian mờ để dự báo đạt kết quả chính xác hơn.

Chen [2] đã đưa ra phương pháp mới đơn giản và hữu hiệu hơn so với phương pháp của Song và Chissom bằng cách sử dụng các phép tính số học thay vì các phép tính hợp max-min phức tạp trong xử lý mối quan hệ mờ. Phương pháp của Chen cho hiệu quả cao hơn về mặt sai số dự báo và độ phức tạp của thuật toán. Nhiều công trình tiếp theo đã sử dụng cách tiếp cận này để dự báo cho chuỗi thời gian. Huang đã sử dụng các thông tin có trước trong tính chất của chuỗi thời gian như mức độ tăng giảm để đưa ra mô hình heuristic chuỗi thời gian mờ.

Mô hình chuỗi thời gian mờ đang có nhiều ứng dụng trong công tác dự báo. Tuy nhiên kết quả dự báo của các phương pháp đề xuất còn chưa cao. Do đó việc tìm tòi các mô hình có độ chính xác cao hơn và thuật toán đơn giản hơn đang là một ưu tiên.

Để nâng cao hiệu quả và độ chính xác của thuật toán, trong những năm gần đây đã có hàng loạt công trình đưa ra nhiều kỹ thuật khác nhau. Những công cụ trong lý thuyết tính toán mềm, khai phá dữ liệu, mạng nơ ron và các giải thuật tiến hoá đều được đưa vào sử dụng. Một số tác giả sử dụng phương pháp phân cụm như công trình của Chen et al, tập thô hay sử dụng khái niệm tối ưu đám đông để xây dựng các thuật toán trong mô hình chuỗi thời gian mờ. Ngoài ra, một số tác giả khác đã sử dụng thêm thông tin khác trong chứng khoán để dự báo chính xác hơn các chỉ số chứng khoán. Từ đó nảy sinh ra mô hình chuỗi thời gian mờ loại 2 khi đồng thời với chuỗi thời gian chính còn sử dụng số liệu của các tham số phụ để đưa ra dự báo.

Một trong các hướng được phát triển là sử dụng mối quan hệ mờ bậc cao trong mô hình chuỗi thời gian mờ. Chen [3] tiếp tục là người đi đầu khi xây dựng được thuật toán để xử lý mối quan hệ mờ bậc cao. Sau đó hướng này được một số tác giả khác tiếp cận và ứng dụng trong các công trình của mình.

Trong những năm gần đây một số công trình đã được hoàn thành theo hướng nâng cao độ chính xác và giảm khối lượng tính toán trong mô hình chuỗi thời gian mờ như các công trình của Chen và Hsu, Huarng, Singh,... Một cách tiếp cận khác cho mô hình chuỗi thời gian mờ là sử dụng những kỹ thuật khác trong khai phá dữ liệu như phân cụm, mạng nơ ron, giải thuật di truyền hay tối ưu đám đông ... để xây dựng mô hình và làm tăng tính hiệu quả của thuật toán.

Dự báo chuỗi thời gian sử dụng mô hình chuỗi thời gian mờ có một số bước cơ bản như sau: Xác định tập nền, Phân chia tập nền thành các khoảng, Mờ hoá các giá trị lịch sử, Xác định các mối quan hệ mờ, Dự báo và cuối cùng là giải mờ. Nhiều nhà khoa học đã cho thấy cách phân chia khoảng có ảnh hưởng rất lớn đến độ chính xác của thuật toán. Nếu phân các khoảng có độ dài lớn thì số phép tính giảm nhưng sẽ có sự phân tán kết quả, còn nếu chia khoảng nhỏ mất ý nghĩa của dự báo. Các tác giả có đề xuất nhiều cách khác nhau để phân khoảng như chia ngẫu nhiên, dựa vào giá trị trung bình, dựa vào phân bố hay dựa vào mật độ phân bố. Mỗi phương pháp được sử dụng trong các trường hợp khác nhau và đều cho kết quả tốt hơn so với phương pháp truyền thống. Từ đây cũng có thể thấy rõ sự ảnh hưởng của phương pháp chia khoảng đến kết quả dự báo.

Có thể thấy rằng nhiều tác giả đã đưa ra phương pháp nâng cao độ chính xác của mô hình khác nhau nhưng phương pháp cơ bản đầu tiên là các phương pháp phân khoảng. Cần thiết phải có những đánh giá và tổng kết các phép phân chia độ dài khoảng để sử dụng trong nhiều bài toán khác nhau. Đó chính là lý do em đã lựa chọn đề tài “Các phương pháp chia khoảng trong mô hình chuỗi thời gian mờ” làm đề tài cho luận văn tốt nghiệp của mình.

Nội dung chính của luận văn có cấu trúc như sau :

Chương 1: Các kiến thức cơ bản về chuỗi thời gian và tập mờ

Chương 2: Mô hình chuỗi thời gian mờ

Chương 3: Các phương pháp chia khoảng trong mô hình chuỗi thời gian mờ

Luận văn này được hoàn thành dưới sự hướng dẫn tận tình của TS Nguyễn Công Điều, tác giả xin bày tỏ lòng biết ơn chân thành của mình đối với thầy. Tác giả xin chân thành cảm ơn các thầy, cô giáo Viện Công nghệ thông tin, trường Đại học Công nghệ thông tin và Truyền thông đã tham gia giảng dạy, giúp đỡ em trong suốt quá trình học tập nâng cao trình độ kiến thức. Tuy nhiên vì điều kiện thời gian và khả năng có hạn nên luận văn không thể tránh khỏi những thiếu sót. Tác giả kính mong các thầy cô giáo và bạn đóng góp ý kiến để đề tài được hoàn thiện hơn.

CHƯƠNG 1: CÁC KIẾN THỨC CƠ BẢN VỀ CHUỖI THỜI GIAN VÀ TẬP MỜ

Chương 1 giới thiệu các kiến thức cơ bản về chuỗi thời gian và tập mờ, trọng tâm là trình bày về một lớp mô hình chuỗi thời gian hết sức thông dụng trong thực tế. Đó là mô hình quy trình trượt ARMA (Autoregressive Moving Average). Bao gồm các nội dung: đặc trưng của quá trình ARMA, phương pháp ước lượng tham số của lớp mô hình này và hạn chế của nó khi áp dụng với chuỗi thời gian tài chính. Một số kiến thức cơ bản về hệ mờ có liên quan tới mô hình chuỗi thời gian mờ

1.1. Chuỗi thời gian và quá trình ngẫu nhiên

1.1.1. Khái niệm chuỗi thời gian và quá trình ngẫu nhiên

Một chuỗi thời gian là một dãy các giá trị quan sát $X := \{x_1, x_2, \dots, x_n\}$ được xếp thứ tự diễn biến thời gian với x_1 là các giá trị quan sát tại thời điểm đầu tiên, x_2 là quan sát tại thời điểm thứ 2 và x_n là quan sát tại thời điểm thứ n .

Ví dụ: Các báo cáo tài chính mà ta thấy hằng ngày trên báo chí, tivi hay Internet về các chỉ số chứng khoán, tỷ giá tiền tệ, chỉ số tiêu dùng đều là những thể hiện rất thực tế của chuỗi thời gian.

Bước đầu tiên của việc phân tích chuỗi thời gian là chọn một mô hình toán học phù hợp với tập dữ liệu cho trước $X := \{x_1, x_2, \dots, x_n\}$ nào đó. Để có thể nói về bản chất của những quan sát chưa diễn ra, ta giả thiết mỗi quan sát x_t là một giá trị thể hiện của biến ngẫu nhiên X_t với $t \in T$. Ở đây T được gọi là tập chỉ số. Khi đó ta có thể coi tập dữ liệu $X := \{x_1, x_2, \dots, x_n\}$ là thể hiện của quá trình ngẫu nhiên $\{X_t, t \in T\}$. Và vì vậy, ta có thể định nghĩa một quá trình ngẫu nhiên như sau:

Định nghĩa 1.1(Quá trình ngẫu nhiên)

Một quá trình ngẫu nhiên là một họ các biến ngẫu nhiên $\{X_t, t \in T\}$ được định nghĩa trên một không gian xác suất (Ω, A, P) .

Chú ý:

Trong việc phân tích chuỗi thời gian, tập chỉ số T là một tập các thời điểm, ví dụ như là tập $\{1, 2, \dots\}$ hay tập $(-\infty, +\infty)$. Cũng có những quá trình ngẫu nhiên có T không phải