

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

TONG HER

ỨNG DỤNG XÍCH MARKOV
TRONG XÂY DỰNG MÔ HÌNH DỰ ĐOÁN

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên - 2022

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

TONG HER

**ỨNG DỤNG XÍCH MARKOV
TRONG XÂY DỰNG MÔ HÌNH DỰ ĐOÁN**

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 848 01 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: TS. TRẦN QUANG QUÝ

Thái Nguyên - 2022

LỜI CẢM ƠN

Đầu tiên, tôi xin gửi lời cảm ơn sâu sắc nhất tới thầy **TS. Trần Quang Quý**, thầy đã tận tình hướng dẫn và chỉ bảo và giúp đỡ tôi thực hiện luận văn này.

Tôi xin chân thành cảm ơn Ban giám hiệu trường Đại học Công nghệ thông tin và Truyền thông đã quan tâm tới các học viên quốc tế; Các thầy, cô giáo trong trường đã tận tình giảng dạy và tạo điều kiện thuận lợi để tôi có thể học tập và rèn luyện trong suốt thời gian theo học tại Trường.

Tôi xin chân thành cảm ơn những người thân và các bạn bè đã chia sẻ, giúp đỡ tôi hoàn thành luận văn này.

Mặc dù đã hết sức cố gắng, nhưng do thời gian và kinh nghiệm nghiên cứu còn có nhiều hạn chế, tiếng Việt chưa thực sự thông thạo nên luận văn vẫn còn một số thiếu sót. Kính mong các Thầy/Cô và các bạn góp ý để tôi có thể kịp thời chỉnh sửa.

Tôi xin chân thành cảm ơn!

Thái Nguyên, tháng 10 năm 2022

Học viên thực hiện

Tong HER

LỜI CAM ĐOAN

Tôi xin cam đoan nội dung và kết quả nghiên cứu trong luận văn này là trung thực và những nội dung trùng lặp với các đề tài khác đều đã được chú thích tham khảo theo đúng quy định. Tôi cũng xin cam đoan 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.

Thái Nguyên, tháng 10 năm 2022

Tác giả

TONG HER

MỤC LỤC

MỤC LỤC	1
MỤC LỤC HÌNH ẢNH	2
MỤC LỤC BẢNG BIỂU	3
DANH MỤC VIẾT TẮT	4
MỞ ĐẦU	6
CHƯƠNG 1: TỔNG QUAN VỀ XÍCH MARKOV	7
1.1. Tìm hiểu về quá trình Markov ngẫu nhiên	7
1.2. Tìm hiểu về xích Markov với thời gian rời rạc (DTMC)	13
1.3. Các định nghĩa liên quan đến xích Markov	21
CHƯƠNG 2: XÍCH MARKOV RỜI RẠC VÀ THUẬN NHẤT	23
2.1. Ma trận xác suất chuyển vị	23
2.2. Phân phối ban đầu	32
2.3. Ví dụ về xích Markov rời rạc và thuận nhất	33
2.4. Mở rộng: Xích Markov có hữu hạn trạng thái	36
CHƯƠNG 3: ỨNG DỤNG XÍCH MARKOV VÀ XÂY DỰNG MÔ HÌNH DỰ ĐOÁN TÀI CHÍNH	42
3.1. Ứng dụng xích Markov	42
3.1.1. Cấu trúc gói xích Markov trong R.	42
3.1.2. Cú pháp sử dụng phân tích xác suất đối tượng xích Markov	47
3.1.3. Cú pháp, hàm sử dụng để phân tích thời gian rời rạc (DTMC)....	54
3.1.4. Xây dựng chương trình mô phỏng dự đoán thời tiết	56
3.2. Xây dựng mô hình dự đoán tài chính	59
3.2.1. Dự đoán dữ liệu tuần tự.....	59
3.2.2. Ma trận chuyển tiếp.....	61
3.2.3. Triển khai mô hình dự đoán hành vi thị trường chứng khoán	62
3.2.4. Thu thập dữ liệu S&P 500.....	64
3.2.5. Kết quả từ mô hình dự báo.....	70
KẾT LUẬN	72
TÀI LIỆU THAM KHẢO	73

MỤC LỤC HÌNH ẢNH

Hình 1.1. Mô phỏng quá trình ngẫu nhiên	8
Hình 2.1. Các trạng thái của xích Markov được biểu thị	25
Hình 2.2. Trạng thái lặp lại và tạm thời	37
Hình 3.1. Mô phỏng xích Markov	45
Hình 3.2. Nhập và xuất đối với một đối tượng xích Markov	46
Hình 3.3. Xích Markov MathematicaMC	53
Hình 3.4. Dữ liệu mẫu trong huấn luyện mô hình Markov	60
Hình 3.5. Mô hình chuyển trạng thái Markov	60
Hình 3.6. Hiển thị 10 dữ liệu chứng khoán đầu tiên	69
Hình 3.7. Hiển thị 10 dữ liệu chứng khoán cuối cùng (mới nhất)	69
Hình 3.8. Trực quan dữ liệu GSPC	69

MỤC LỤC BẢNG BIỂU

Bảng 1.1. Bảng phân phối xác suất cho $X(0)$	21
Bảng 2.1. Xác suất trạng thái trong 15 thời kỳ	26
Bảng 2.2. Bảng ma trận xác suất chuyển đổi.....	29
Bảng 2.3. Phân phối xác suất tại các siêu thị.....	34
Bảng 2.4. Phân phối xác suất cho $X(1)$	35
Bảng 3.1. Các phương thức để xử lý đối tượng xích Markov	44
Bảng 3.2 Các tập dữ liệu được đính kèm thư viện markovchain.....	47
Bảng 3.3. Các file đính kèm thư viện markovchain mô phỏng	47
Bảng 3.4. Các phương thức trong gói markovchain	48
Bảng 3.5. Các hàm thống kê Markovchain.....	54
Bảng 3.6. Ví dụ về dữ liệu chứng khoán, giá đóng cửa S&P 500	64

DANH MỤC VIẾT TẮT

Chữ viết tắt	Chữ đầy đủ
DTMC	Discrete Time Markov Chain
MLE	Maximum Likelihood Estimation
GIS	Geographic Information System
RS	Remote Sensing
ACF	Autocorrelation Function
ANN	Artificial Neural Network
AIC	Akaike Information Criterion
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayessian Information Criterion
BPNN	Back Propagation Neural Network
BWP	Backward Probabilities
CMC	Comercial Higher Order Markov Chain
DJIA	Dow Jones Industrial Average Index
EM	Expectation-Maximization
FTS	Fuzzy Time Series
FWP	Forward Probabilities
GA	Genetic Algorithm
GDP	Gross Domestic Product
GPS	Global Positioning System
IMC	Improved Higher Order Markov Chain

MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MC	Markov Chain
PCA	Principle Component Analysis
RMSE	Root Mean Square Error
SSE	Shanghai Stock Exchange
SVM	Support Vector Machine

MỞ ĐẦU

Đầu thế kỉ XX, nhà Toán học nổi tiếng người Nga *A. A. Markov* đã đưa ra mô hình Toán học để mô tả chuyển động của các phân tử chất lỏng trong một bình kín. Sau này mô hình được phát triển và mang tên là “Quá trình Markov”. Nhiều mô hình ngẫu nhiên trong Kinh tế, kỹ thuật, dân số học, di truyền học... đều được dựa trên cơ sở là quá trình Markov. Xích Markov là một trường hợp riêng của quá trình Markov khi ta có thể đánh số được các trạng thái.

Xích Markov được ứng dụng rộng rãi làm mô hình thống kê của nhiều quá trình đời thực, như là nghiên cứu hệ thống điều khiển hành trình trong các xe mô tô, hàng đợi hay hàng người đến sân bay chờ làm thủ tục, tỉ giá hối đoái tiền tệ và sự biến đổi của dân số quần thể. Quá trình Markov là cơ sở cho phương pháp mô phỏng ngẫu nhiên xích Markov Monte Carlo, được dùng để mô phỏng việc lấy mẫu từ một phân bố xác suất phức tạp, và có ứng dụng trong thống kê Bayes, nhiệt động lực học, cơ học thống kê, vật lý, hóa học, kinh tế, tài chính, xử lý tín hiệu, lý thuyết thông tin và trí tuệ nhân tạo.

Năm 1998, Lawrence Page, Sergey Brin, Rajiv Motwani và Terry Winograd đã công bố bài báo “The PageRank Citation Ranking: Bringing Order to the Web”, trong bài báo mô tả thuật toán PageRank (hay còn gọi là thuật toán xếp hạng trang) nổi tiếng và hiện nay đã trở thành nền tảng của Google. Chưa đầy hai thập kỷ sau, Google đã trở thành một gã khổng lồ, và cho dù thuật toán của hãng đã phát triển rất nhiều, PageRank vẫn là "biểu tượng" trong các thuật toán xếp hạng của Google. Một trong những nền tảng quan trọng của PageRank chính là xích Markov.

Trong nội dung của luận văn này trình bày các khái niệm chính về quá trình ngẫu nhiên rời rạc, xích Markov và ứng dụng của xích Markov trong việc xây dựng mô hình dự đoán trong thực tế.