

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN KHOA HỌC MÁY TÍNH

TRẦN NGỌC TRUNG-0312333

XÂY DỰNG ỨNG DỤNG HỖ TRỢ DỰ ĐOÁN CHỨNG KHOÁN

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

GIÁO VIÊN HƯỚNG DẪN
PGS.TS LÊ HOÀI BẮC

KHÓA 2003-2007

[illegible]

2

This image shows a full page of a notebook or worksheet. It features approximately 20 horizontal rows. Each row is defined by two parallel dotted lines, creating a series of uniform spaces for handwriting practice. The paper is otherwise completely blank, with no margins, text, or other markings.

TpHCM, ngày tháng năm
Giáo viên phản biện

LỜI CẢM ƠN

Em xin chân thành cảm ơn khoa Công Nghệ Thông Tin, trường Đại Học Khoa Học Tự Nhiên đã hỗ trợ tạo nhiều điều kiện thuận lợi cho chúng em trong quá trình học tập cũng như thực hiện khóa luận tốt nghiệp.

Em xin chân thành cảm ơn cô Phan Thị Bích Nguyệt đã hỗ trợ, định hướng và tạo điều kiện cho em có dịp tìm hiểu những vấn đề nghiệp vụ chứng khoán.

Xin cảm ơn bạn Nguyễn Quốc Anh đã có nhiều đóng góp ý kiến về nghiệp vụ trong quá trình thực hiện khóa luận và giúp đỡ về vấn đề dữ liệu liên quan đến nghiệp vụ.

Em xin chân thành cảm ơn quý Thầy Cô trong Khoa CNTT đã tận tình giảng dạy, trang bị cho em những kiến thức cần thiết trong suốt quá trình học tập và nghiên cứu tại Khoa.

Con xin ghi nhớ công ơn của ba mẹ đã sinh thành, nuôi dưỡng và dạy dỗ con trưởng thành như ngày hôm nay. Ba mẹ luôn là chỗ dựa tinh thần hết sức lớn lao những khi con vấp ngã và gặp những khó khăn trong cuộc sống.

Xin gửi lời cảm ơn đến các bạn bè đã gửi lời hỏi thăm và động viên cũng như giúp đỡ về tài liệu trong quá trình thực hiện khóa luận.

Cuối cùng em xin gửi lời cảm ơn sâu sắc đến thầy Lê Hoài Bắc đã tận tình hướng dẫn, truyền đạt cho em những kiến thức, những lời khuyên quý báu. Cùng với lời động viên nhắc nhở và khích lệ của thầy trong lúc em gặp những khó khăn, trở ngại khi thực hiện khóa luận.

Mặc dù đã nỗ lực hết sức để hoàn thành tốt khóa luận nhưng cũng không thể tránh khỏi những thiếu sót, kính mong các Thầy Cô thông cảm và tận tình chỉ bảo.

TP. Hồ Chí Minh 7/2007

Sinh viên thực hiện

Trần Ngọc Trung

ĐỀ CƯƠNG CHI TIẾT

Tên Đề tài: Xây dựng ứng dụng hỗ trợ dự đoán chứng khoán
Giáo viên hướng dẫn: PGS.TS. Lê Hoài Bắc
Thời gian thực hiện: 12/11/2006 – 16/07/2007
Sinh viên thực hiện: Trần Ngọc Trung - 0312333
Loại đề tài: Khóa luận sử dụng một số kỹ thuật trong tính toán thông minh, nhằm xây dựng các hệ thống đánh giá và dự đoán chứng khoán trên thị trường giao dịch chứng khoán TPHCM. Chi tiết hơn là sử dụng mô hình CBR cho việc đánh giá giá trị cổ phiếu, và sử dụng một số các phương pháp phân tích kỹ thuật cho dự đoán giá cổ phiếu trong tương lai.

Nội Dung Đề Tài: Trong dự đoán chứng khoán có 2 hướng: (1) Phân tích cơ bản, (2) Phân tích kỹ thuật. Khóa luận sẽ tiếp cận trên cả 2 hướng. Với mong muốn hỗ trợ một phần nào đó sự giảm thiểu rủi ro cho các nhà đầu tư chứng khoán. Đây là một trong những định hướng áp dụng tin học vào giải quyết các bài toán kinh tế nói chung và dự đoán chứng khoán nói riêng. <i>Trong phân tích cơ bản:</i> Kỹ thuật được áp dụng là mô hình CBR. Một tình huống trong CBR là dựa trên nhiều thông số thường được sử dụng trong phân tích cơ bản. Trong hệ CBR sẽ kết hợp một số những phương pháp khác trong tính toán mềm là logic mờ và mạng nơron. <i>Trong phân tích kỹ thuật:</i> Chủ yếu là dựa trên bài toán dự đoán giá cổ phiếu trong tương
--

lai. Vấn đề này được giải quyết bằng cách đề xuất một số những thuật toán khác nhau:

- Mạng nơron trong dự đoán giá đóng cửa.
 - Dựa trên chỉ số giá đóng cửa.
 - Dựa trên các thông số kỹ thuật.
- Mô hình ECBR
- Mô hình Fuzzy time-series kết hợp Fibonacci

Hai hướng đều đạt được một số những kết quả nhất định, mặc dù không phải tất cả các phương pháp đề xuất đều thực sự hiệu quả, có những phương pháp tốt, nhưng có những phương pháp kết quả chưa thực sự thuyết phục. Tuy nhiên, đó cũng là những đề xuất ban đầu, những hướng đi ban đầu cho bản thân trong lĩnh vực liên quan đến các bài toán kinh tế nói chung và dự đoán chứng khoán nói riêng.

- Phương pháp CBR: Với 300 tình huống trong cơ sở tình huống ban đầu, và 50 tình huống được sử dụng trong quá trình kiểm chứng, kết quả đạt được là 88%.
- Trong 4 phương pháp đề xuất trong phân tích kỹ thuật có một số kết quả:
 - Phương pháp mạng nơron sử dụng giá đóng cửa: Mức độ dự đoán lỗi trung bình ở mức độ chấp nhận được. Nó là phương pháp có chỉ số lỗi thấp thứ 2.
 - Phương pháp mạng nơron sử dụng các thông số kỹ thuật: Phương pháp này đạt được kết quả không tốt. Mặc dù trên ý tưởng được hình thành khá hợp lý.
 - Phương pháp ECBR: Tương tự như mạng nơron sử dụng giá đóng cửa thì phương pháp này cũng có hệ số thấp hơn một chút. Tuy nhiên một yếu điểm rất lớn của phương pháp này là tốc độ thực hiện thuật toán là rất chậm.
 - Phương pháp Fibonacci: Đây là phương pháp tốt nhất trong tất cả các phương pháp. Một trong những vấn đề khá quan trọng mà cần đề cập ở đây là so với

một số bài báo tìm hiểu về lĩnh vực liên quan thì thuật toán có hệ số lỗi tốt hơn. Nhưng cũng chưa thể đánh giá gì vì dữ liệu được sử dụng để đánh giá là khác nhau.

Trên đây là một số những giới thiệu một cách khái quát về những vấn đề đang và sẽ được tiếp tục phát triển sau này.

Các thuật toán hiệu thời đều chỉ là những thuật toán dự đoán giá đóng cửa của ngày hôm sau thực sự với giá dự đoán ngày hôm sau chưa giúp được gì nhiều cho nhà đầu tư trong vấn đề dự đoán chứng khoán. Vì vậy một trong những hướng tiên quyết sẽ được phát triển trong tương lai sẽ là tăng khoảng thời gian dự đoán lên từ 5 – 10 ngày.

Kế Hoạch Thực Hiện: Để bắt đầu thực hiện luận văn cần có một kế hoạch rõ ràng với mục tiêu có thể hoàn thành luận văn với kết quả tốt nhất. Dưới đây là các mốc thời gian chi tiết cho quá trình thực hiện luận văn.

Giai đoạn	Khoảng thời gian	Công việc	Kết quả mong muốn
1	12/11/2006 22/11/2006	Tìm các tài liệu liên quan đến các vấn đề trong luận văn cần thực hiện.	Có được những tài liệu cần thiết trong quá trình thực hiện luận văn.
2	23/11/2006 23/12/2006	Tìm hiểu các vấn đề nền tảng liên quan đến những vấn đề cần đặt ra trong luận văn. <ul style="list-style-type: none"> • Tìm hiểu về CBR • Tìm hiểu về nghiệp 	Có được những kiến thức tổng quát, hiểu được những khái niệm và những vấn đề cũng như định hướng đi trong luận văn.

		vụ chứng khoán.	
	24/12/2006 06/01/2007	Định hướng đi, tìm hiểu các phương pháp, đề xuất những ý tưởng.	Có được hướng đi đúng đắn trong xây dựng các ứng dụng trong xây dựng các bài toán hỗ trợ trong đầu tư chứng khoán.
	06/01/2007 20/06/2007	Bắt tay vào xây dựng ứng dụng thực tế, đồng thời với việc xây dựng báo cáo luận văn.	Xây dựng hoàn chỉnh ứng dụng và hoàn tất báo cáo cho khóa luận tốt nghiệp
Xác nhận của GVHD		Ngày.....tháng.....năm..... SV Thực hiện	

LỜI MỞ ĐẦU

Thị trường chứng khoán là loại thị trường bậc cao của nền kinh tế thị trường ở hầu hết các nước trên thế giới, đặc biệt với các nước có ngành công nghiệp phát triển đều có thị trường chứng khoán có trình độ phát triển cao và qui mô rất lớn.

Ở Việt Nam, trong xu thế hội nhập quốc tế, chúng ta đã hình thành thị trường chứng khoán, thông qua thực tiễn hoạt động, thị trường chứng khoán Việt Nam cũng đã chứng minh sự cần thiết và vai trò to lớn của nó.

Đã hoạt động được 7 năm, kể từ khi ra đời năm 2000. Thị trường đã có những giai đoạn thăng trầm, không ổn định. Điều này thường xảy ra với các thị trường chứng khoán non trẻ như Việt Nam. Tuy nhiên, nhiều chuyên gia nước ngoài đã khẳng định tiềm lực khổng lồ chưa được khai thác trong thị trường chứng khoán Việt Nam và thực sự thời gian gần đây nó đã tự ổn định dần.

Sự ra đời và phát triển của thị trường chứng khoán, kéo theo phong trào đầu tư chứng khoán trong người dân phát triển mạnh. Ngày càng có nhiều người, nhiều đối tượng xã hội quan tâm. Tuy nhiên, đi cùng với lợi nhuận hấp dẫn mang lại từ thị trường chứng khoán là mức độ rủi ro trong đầu tư chứng khoán. Có nhiều nhà đầu tư nhưng lại không nhiều người thực sự có kiến thức đầy đủ cũng như được đào tạo một cách bài bản để đầu tư thật sự hiệu quả. Họ chỉ chạy theo cơn sốt đầu tư và bị cuốn bởi sự hấp dẫn của nó.

Với mong muốn giúp cho các nhà đầu tư, đặc biệt là các nhà đầu tư chưa thực sự có nhiều kinh nghiệm có thể giảm thiểu được một mức độ rủi ro trong đầu tư và hỗ trợ trong quá trình ra quyết định. Mục tiêu trong khóa luận là xây dựng một số phương pháp trong việc đánh giá cũng như đưa ra những dự đoán về giá cả chứng khoán trong tương lai.

Khóa luận “Xây dựng ứng dụng hỗ trợ dự đoán chứng khoán” sẽ trình bày cả hai hướng khác nhau trong đầu tư chứng khoán “Phân tích cơ bản” và “Phân tích kỹ thuật”.

Trong đó, sẽ tập trung khá nhiều vào lĩnh vực “Phân tích kỹ thuật” với mục tiêu là có thể dự đoán được giá cả thị trường của các cổ phiếu trong tương lai.

Phần trình bày gồm 5 chương:

TỔNG QUAN: chương mở đầu của khóa luận, trình bày tổng quan, lý do thực hiện và những mục tiêu và hướng tiếp cận của khóa luận cũng như các đối tượng nghiên cứu.

KIẾN THỨC NỀN TẢNG: trình bày một số những kiến thức cơ bản được sử dụng trong khóa luận.

CÁC PHƯƠNG PHÁP VÀ MÔ HÌNH DỰ ĐOÁN: Xây dựng các mô hình đánh giá và đề xuất các thuật toán dự đoán giá tương lai. Các phương pháp sẽ được trình bày một cách chi tiết từng bước.

KẾT QUẢ VÀ ĐÁNH GIÁ: Những kết quả đạt được ở những mô hình và các phương pháp dự đoán. Đồng thời đưa ra những nhận xét đánh giá và so sánh các phương pháp này.

TỔNG KẾT: Tóm tắt những vấn đề đã được trình bày, những vấn đề đã giải quyết, những vấn đề chưa giải quyết được và trình bày một số hướng phát triển của khóa luận.

CẤU TRÚC LUẬN VĂN

Chương 1.	TỔNG QUAN.....	17
1.1	Giới thiệu khái quát	17
1.2	Lý do thực hiện khóa luận	18
1.3	Mục tiêu khóa luận	18
1.4	Hướng tiếp cận và các đối tượng nghiên cứu	19
1.4.1.	Hướng tiếp cận	19
1.4.2.	Đối tượng nghiên cứu.....	20
Chương 2.	CÁC KIẾN THỨC NỀN TẢNG	22
2.1	Thị trường chứng khoán	22
2.1.1.	Phân tích cơ bản	23
2.1.2.	Phân tích kỹ thuật.....	24
2.2	Suy diễn dựa vào tình huống - CBR (Case-Based Reasoning).....	25
2.2.1.	Giới thiệu	25
2.2.2.	Các thành phần của CBR	26
2.2.2.1.	Biểu diễn, đánh chỉ số tình huống (Case representation and indexing)..	28
2.2.2.2.	Rút trích tình huống (Case Retrieval)	30
2.2.2.3.	Sự thích nghi tình huống (Case adaptation).....	31
2.2.2.4.	Bảo trì tình huống.....	34
Chương 3.	CÁC PHƯƠNG PHÁP VÀ MÔ HÌNH DỰ ĐOÁN	35
3.1	Áp dụng trong phân tích cơ bản.....	35
3.1.1.	Đánh giá cổ phiếu dựa trên phân tích cơ bản.....	35
3.1.1.1.1.	Chức năng và định hướng của phương pháp	36
3.1.1.1.2.	Biểu diễn tình huống.....	36
3.1.1.1.3.	Hệ thống làm mờ (Fuzzifier)	40
3.1.1.1.4.	Đánh chỉ số tình huống	44
3.1.1.1.5.	Rút trích tình huống.....	45
3.1.1.1.6.	Thích nghi tình huống.....	49
3.1.1.1.7.	Bảo trì cơ sở tình huống.....	50
3.2	Áp dụng trong phân tích kỹ thuật	51
3.2.1.	Mô hình mạng nơron	51
3.2.1.1.	Dự đoán dùng mô hình ARIMA	53
3.2.1.1.1.	Tiền xử lý dữ liệu (Data preprocessing)	55
3.2.1.1.2.	Biến đổi dữ liệu (Data Scaling)	56
3.2.1.1.3.	Cấu trúc mạng nơron	57
3.2.1.1.4.	Huấn luyện mạng	59
3.2.1.1.5.	Phương thức đánh giá dự đoán	61
3.2.1.2.	Dự đoán dùng các chỉ số kỹ thuật.....	62
3.2.2.	Mô hình ECBR	64
3.2.2.1.	Thể hiện case và đánh chỉ số case.....	64
3.2.2.2.	Đánh chỉ số tình huống	65
3.2.2.3.	Giải quyết vấn đề với CBR truyền thống	67
3.2.2.4.	Giới thiệu phương pháp Ensemble CBR	67

3.2.2.5.	Thuật toán Ensemble CBR	68
3.2.2.6.	Áp dụng thuật toán đề xuất vào bài toán dự đoán chứng khoán.....	70
3.2.3.	Mô hình ứng dụng Fibonacci trong lý thuyết sóng Elliott.....	74
3.2.3.1.	Chuỗi thời gian mờ (Fuzzy time-series) và lý thuyết sóng Elliot.....	75
3.2.3.2.	Mô hình chuỗi thời gian mờ (Fuzzy time-series) chuẩn	76
3.2.3.3.	Mô hình đề xuất	78
Chương 4.	KẾT QUẢ VÀ ĐÁNH GIÁ.....	87
4.1	Phương pháp CBR cho đánh giá chứng khoán	87
4.1.1.	Dữ liệu đánh giá	87
4.1.2.	Một số kết quả đạt được	87
4.1.3.	Đánh giá.....	90
4.1.3.1.	Ưu điểm	90
4.1.3.2.	Nhược điểm	91
4.1.4.	Đề xuất phương pháp cải tiến trong tương lai	91
4.2	Các phương pháp đạt được trong dự đoán kỹ thuật.....	91
4.2.1.	Phương pháp Random Walk	95
4.2.2.	Phương pháp mạng nơron	99
4.2.2.1.	Kết quả đạt được	99
4.2.2.2.	Đánh giá thuật toán	101
4.2.2.3.	Đề xuất phương pháp cải tiến trong tương lai	102
4.2.3.	Sử dụng các thông số kỹ thuật trong mạng nơron.....	103
4.2.3.1.	Một số kết quả đạt được	103
4.2.3.2.	Đề xuất phương pháp cải tiến trong tương lai	105
4.2.4.	Phương pháp ECBR	105
4.2.4.1.	Một số kết quả đạt được	105
4.2.4.2.	Đánh giá phương pháp	106
4.2.4.3.	Đề xuất thuật toán cải tiến trong tương lai	107
4.2.5.	Phương pháp Fibonacci-Elliott	107
4.2.5.1.	Một số kết quả đạt được	107
4.2.5.2.	Đánh giá thuật toán	111
4.2.6.	So sánh kết quả đạt được của một số thuật toán trong phân tích kỹ thuật. ...	111
4.2.7.	So sánh với một số công trình khác	115
Chương 5.	KẾT LUẬN.....	116
	TÀI LIỆU THAM KHẢO	118
	PHỤ LỤC	122
	➤ Mô hình ARIMA	122
	➤ Một số những chỉ số kỹ thuật.....	125
	➤ Lý thuyết sóng Elliott và chuỗi Fibonacci	132
	Fibonacci trong phân tích kỹ thuật - Dự đoán xu hướng giá chứng khoán	132
	Kết hợp Fibonacci và sóng Elliott dự đoán xu hướng thị trường	137

DANH SÁCH CÁC HÌNH VÀ CÁC BẢNG

Danh sách các hình

Hình 1 : Một phần bảng giao dịch trực tuyến TPHCM 13/07/2007	23
Hình 2 : Các kiến thức liên quan đến phân tích kỹ thuật.....	24
Hình 3 : Một biểu đồ thể hiện phân tích kỹ thuật chỉ số VNINDEX	25
Hình 4 : Chu kì CBR.....	26
Hình 5 : Các thành phần cơ bản của CBR.....	28
Hình 6 : Quá trình rút trích tình huống	31
Hình 7 : Quá trình rút trích tình huống	32
Hình 8 : Quá trình thích ứng tình huống.....	33
Hình 9 : Hệ thống mờ.....	41
Hình 10 : Hàm mờ trapzoid.....	41
Hình 11 : Sơ đồ mô hình CBR cho đánh giá cổ phiếu.....	43
Hình 12 : Một ví dụ trong phương pháp đánh chỉ số.....	45
Hình 13 : Ví dụ về phương pháp mờ hóa thuộc tính EPS sử dụng cho đầu vào mạng nơron	47
Hình 14 : Bộ làm mờ biến ngôn ngữ.....	48
Hình 15 : Cấu trúc mạng nơron nhằm phân loại tình huống sau khi làm mờ	48
Hình 16 : Mô hình thích nghi tình huống.....	49
Hình 17 : Giá giao dịch cơ phiếu REE (05/02/2007 – 05/05/2007)	54
Hình 18 : Giá giao dịch cơ phiếu FPT (05/02/2007 – 05/05/2007)	54
Hình 19 : Giá giao dịch cơ phiếu SAM(05/02/2007 – 05/05/2007)	55
Hình 20 : Mô hình mạng nơron sử dụng ARIMA	58
Hình 21 : Hàm sigmoid	59
Hình 22 : Phương pháp áp dụng chuỗi giá vào mạng nơron.....	60
Hình 23 : Sơ đồ chuẩn hóa ECBR	72
Hình 24 : Lược đồ thuật toán ECBR.....	74
Hình 25 : Mô hình chuỗi thời gian mờ (Fuzzy time series).....	78
Hình 26 : Dự đoán giá KDC sử dụng mạng nơron (25/01/2006 - 12/07/2007)	92
Hình 27 : Dự đoán giá cổ phiếu BBC bằng mạng nơron (01/01/2005 - 12/07/2007)	92
Hình 28 : Dự đoán cổ phiếu BMC bằng ECBR (01/06/2007 - 01/07/2007).....	93
Hình 29 : Dự đoán giá cổ phiếu STB bằng phương pháp Fibonacci (12/07/2006 - 12/07/2007).....	93
Hình 30 : So sánh giữa mô hình RW và Fibonacci.....	94
Hình 31 : Random Walk trên cổ phiếu ABT (01/01/2006 - 12-07-2007)	96
Hình 32 : Biểu đồ RW cổ phiếu RAL (12/2006 - 6/2007).....	97
Hình 33 : Dự đoán REE bằng mạng nơron (3/2006 – 4/2007)	99
Hình 34 : Dự đoán cổ phiếu REE bằng mạng nơron (1/2005 – 1/2006)	100

Hình 35 : Dự đoán cổ phiếu FPT bằng ECBR (01/06/2007 - 01/07/2007).....	106
Hình 36 : Dự đoán cổ phiếu SAM bằng Fibonacci (1/2004 – 4/2004).....	108
Hình 37 : Dự đoán cổ phiếu CAN bằng Fibonacci (2/2006 – 4/2006).....	109
Hình 38 : Biểu đồ so sánh lỗi RMSE của 3 phương pháp.....	112
Hình 39 : Biểu đồ so sánh lỗi MAPE của 3 phương pháp.....	112
Hình 40 : Biểu đồ so sánh lỗi MAE của 3 phương pháp.....	113
Hình 41 : Bảng kết quả trên cổ phiếu REE.....	114
Hình 42 : Bảng kết quả trên cổ phiếu SAM.....	114
Hình 43 : Bảng kết quả trên cổ phiếu SGC.....	115
Hình 44 : So sánh với công trình của Pao.....	115
Hình 45 : Đồ thị biểu diễn A/D giảm trong khi giá tăng.....	126
Hình 46 : Một ví dụ và giá theo hướng của RSI.....	128
Hình 47 : Biểu đồ thể hiện %R.....	129
Hình 48 : Biểu đồ ví dụ về PVT.....	131
Hình 49 : Biểu đồ Fibonacci Arcs.....	134
Hình 50 : Biểu đồ Fibonacci Fans.....	135
Hình 51 : Biểu đồ Fibonacci Retracements.....	136
Hình 52 : Biểu đồ sóng Elliott và các số Fibonacci.....	139
Hình 53 : Sóng Elliott cơ bản.....	143

Danh sách các bảng

Bảng 1 : Biểu diễn tình huống trong phân tích đánh giá chứng khoán.....	37
Bảng 2 : Bảng giá trị mờ hóa các thuộc tính.....	42
Bảng 3 : Phương pháp rút trích k-NN.....	46
Bảng 4 : Dữ liệu ứng dụng cho mạng nơron trong phân tích kĩ thuật.....	62
Bảng 5 : Tình huống trong mô hình ECBR.....	65
Bảng 6 : Bảng đánh chỉ số trong ECBR.....	66
Bảng 7 : Thuật toán CBR truyền thống trong dự đoán chứng khoán[39].....	67
Bảng 8 : Thuật toán ECBR.....	70
Bảng 9 : Công thức chuyển đổi Forward Transform.....	73
Bảng 10 : Công thức chuyển đổi Backward Transform.....	73
Bảng 11 : Bảng kết quả đạt được trong mô hình CBR đánh giá cổ phiếu.....	88
Bảng 12 : Cổ phiếu đại diện cho mỗi nhóm ngành được niêm yết trên thị trường cho việc so sánh các phương pháp.....	95
Bảng 13 : Bảng kết quả dự đoán bằng RW trên một số cổ phiếu đại diện.....	98
Bảng 14 : Bảng kết quả dự đoán bằng mạng nơron trên một số cổ phiếu đại diện..	101
Bảng 15 : Bảng tính hệ số lỗi của phương pháp Fibonacci.....	110
Bảng 16 : Bảng thể hiện tỉ lệ trong chuỗi Fibonacci.....	141

TÓM TẮT KHÓA LUẬN

Khóa luận tìm hiểu về các kỹ thuật áp dụng các phương thức của tính toán thông minh vào các lĩnh vực tài chính, và trong trường hợp này là hỗ trợ đánh giá và dự đoán chứng khoán.

Khóa luận sẽ đi vào hai lĩnh vực cơ bản trong đầu tư chứng khoán là “Phân tích cơ bản” và “Phân tích kỹ thuật”, trong đó sẽ đặt nặng trong “Phân tích kỹ thuật” hay nói rõ hơn là dự đoán sự biến động giá của từng loại cổ phiếu trên thị trường.

Trong phân tích cơ bản sẽ đi vào xây dựng một phương pháp đánh giá giá trị của một cổ phiếu, tức là đánh giá một cổ phiếu hiện tại có triển vọng sau này hay không? Phương pháp này tiếp cận dựa trên một số các chỉ số cơ bản thường được sử dụng trong thị trường chứng khoán Việt Nam cũng như các thị trường chứng khoán trên thế giới. Kỹ thuật áp dụng cho hướng giải quyết này là phương pháp CBR, trong đó có kết hợp logic mờ và mạng nơron nhằm hỗ trợ hiệu quả hơn việc giải quyết vấn đề. Tuy những kết quả đạt được của phương pháp này cũng chỉ dừng lại ở mức độ chấp nhận, nhưng nó là một định hướng một phương pháp theo đánh giá của bản thân là có hiệu quả. Nếu có nhiều thời gian hơn nữa, hi vọng sẽ có những kết quả tốt hơn.

Đối với phân tích kỹ thuật có 4 phương pháp được đề xuất ở đây:

- Phương pháp mạng nơron sử dụng giá đóng cửa.
- Phương pháp mạng nơron sử dụng các chỉ số kỹ thuật.
- Phương pháp ECBR
- Phương pháp Fuzzy time-series Fibonacci.

Với 2 phương pháp của mạng nơron trong toàn bộ khóa luận sẽ sử dụng mạng lan truyền ngược 3 lớp, với luật học delta. Trong đó mạng sử dụng ARIMA áp dụng cho giá đóng cửa định hướng dự đoán giá cổ phiếu ở giá đóng cửa của ngày hôm sau, trong khi phương pháp sử dụng các chỉ số kỹ thuật lại dự đoán 4 ngày tiếp theo.

Phương pháp ECBR có sự áp dụng khác trong dự đoán bằng cách sử dụng nhiều chỉ số hơn là việc chỉ sử dụng một chỉ số giống như mạng nơron. Đây là phương pháp phát triển dựa trên mô hình CBR truyền thống, bằng cách sử dụng “nhiều CBR kết hợp” với các độ đo và các ngữ cảnh khác nhau, nhằm dự đoán giá đóng cửa cuối cùng vào ngày hôm sau.

Phương pháp Fibonacci là một kỹ thuật hoàn toàn khác so với các phương pháp trước đó, trong đó có việc ước lượng một số giá trị cho phương trình Fibonacci nhằm đạt được hệ số lỗi nhỏ nhất có thể.

Trong các phương pháp, có phương pháp không đạt được kết quả như mong muốn. Có những phương pháp có kết quả khá lạc quan. Tuy nhiên, cái cuối cùng đạt được đó là những kinh nghiệm những kiến thức mới trong quá trình thực hiện luận văn.

Tuy nhiên, vì đây cũng là những định hướng ban đầu, những bước đi đầu tiên khi tiếp cận một lĩnh vực tuy không mới trên thế giới nhưng khá mới ở Việt Nam “dự đoán chứng khoán”. Ở đây, cái quan trọng nhất đạt được khi thực hiện xong đề tài không phải quá đặt nặng vấn đề kết quả. Mà cái đạt được là một định hướng đúng sau này, với hy vọng những người khác khi làm về cùng một lĩnh vực sẽ không phải bỡ ngỡ, lúng túng bởi không biết nên bắt đầu tiếp cận lĩnh vực này như thế nào!

Chương 1. TỔNG QUAN

1.1 Giới thiệu khái quát

Thị trường chứng khoán là một thị trường kinh tế, thương mại rất năng động và có tầm quan trọng trong nền kinh tế thế giới. Trong các nước phát triển nói chung, chỉ số của thị trường chứng khoán còn được xem như là một trong những số đo chỉ số phát triển của nền kinh tế.

Đã 7 năm từ ngày ra đời, thị trường chứng khoán Việt Nam đã có những tăng trưởng đáng kể, sự phát triển mạnh mẽ của thị trường chứng khoán và lợi nhuận từ thị trường đem lại đã hấp dẫn ngày càng nhiều người dân, nhiều đối tượng xã hội tham gia. Hiện nay, phong trào đầu tư chứng khoán thực sự đã phát triển mạnh mẽ không những giới công chức nhà nước, những nhà kinh doanh mà ngay cả những người dân bình thường, những học sinh, sinh viên cũng đầu tư vào thị trường hấp dẫn này. Tuy nhiên, ngoài lợi nhuận cao đem lại từ thị trường chứng khoán thì đi kèm với nó là những rủi ro, nhất là những người chưa có kinh nghiệm cũng như chưa được đào tạo bài bản để có thể đầu tư chứng khoán một cách hiệu quả.

Trên thế giới, ở các nước có nền kinh tế thị trường phát triển nói chung và thị trường chứng khoán nói riêng có rất nhiều các ứng dụng phần mềm hỗ trợ việc phân tích, dự đoán và hỗ trợ quyết định. Họ đã đạt được một số kết quả nhất định và hỗ trợ rất nhiều cho những nhà đầu tư nhằm có thể giúp họ có được những quyết định hợp lý cũng như có những chiến lược thích ứng nhằm giảm thiểu mức độ rủi ro thấp nhất có thể. Tuy nhiên theo một số nghiên cứu đã tìm hiểu thì hệ số lỗi đạt được trong dự đoán vẫn khá cao, vẫn chưa thật sự có được những tin tưởng từ những người sử dụng. Nhiều những nghiên cứu tìm hiểu được vẫn chỉ ở mức độ chấp nhận và mang nặng tính lý thuyết.

Đối với thị trường Việt Nam, một thị trường khá trẻ, cũng chưa có nhiều nhà đầu tư thực sự có nhiều vốn kiến thức và am hiểu thị trường. Các nhà đầu tư nghiệp dư tham gia đầu tư với mức độ rủi ro rất lớn khi chưa thực sự có một kiến thức vững vàng, cũng

như chưa có những chiến lược hợp lý. Thị trường càng về sau này sẽ luôn tự điều chỉnh và phát triển. Vì vậy, có hai việc cấp thiết cần cho nhà đầu tư: (1) Cần tìm hiểu những kiến thức về đầu tư chứng khoán (2) Cần có những ứng dụng hỗ trợ họ trong quyết định đầu tư.

1.2 Lý do thực hiện khóa luận

Lý do 1. Xây dựng một ứng dụng có thể hỗ trợ những nhà đầu tư, đặc biệt là những người chưa có nhiều kinh nghiệm có giảm mức độ rủi ro có thể trong đầu tư, cũng như giúp họ có được những cái nhìn rõ ràng trong sự biến đổi giá cả thị trường đặc biệt là các cổ phiếu. Một sự dự đoán đạt được ở mức độ nhất định sẽ là những định hướng tốt cho những nhà đầu tư có thể đưa ra những chiến lược hợp lý cũng như những phương pháp lựa chọn đầu tư hiệu quả.

Lý do 2. 7 năm hoạt động của thị trường là thời gian khá ngắn, do đó thực sự chưa nhiều những người nghiên cứu về lĩnh vực này. Lĩnh vực xây dựng những thuật toán trong tin học nhằm giải quyết những bài toán kinh tế, đặc biệt là dự đoán giá, đánh giá cổ phiếu trong thị trường chứng khoán. Và cũng chưa tìm thấy một nghiên cứu nào ở Việt Nam thực sự đạt được những kết quả đáng kể được công bố, một phần cũng bởi vì thị trường chứng khoán là khá mới mẻ đối với nhiều người. Đây cũng là lý do và động lực để đi đến xây dựng khóa luận.

Lý do 3. Trong quá trình thực hiện khóa luận, mong muốn lớn nhất là xây dựng một thuật toán hiệu quả ở mức độ chấp nhận, và có thể áp dụng ít nhất là cho bản thân khi tham gia đầu tư thực sự trong thị trường chứng khoán sau này.

1.3 Mục tiêu khóa luận

Như đã trình bày ở trên, mục tiêu cuối cùng là xây dựng một ứng dụng có thể giúp đỡ được một phần cho các nhà đầu tư với mức độ tin cậy chấp nhận được.

Từ lý thuyết đến thực tế là một quá trình rất dài và không đơn giản, vì vậy từ những tìm hiểu lý thuyết để có thể đưa ra thực tế áp dụng là cả một vấn đề khó khăn. Tuy vậy đó cũng chính là mục tiêu trước mắt trong quá trình thực hiện khóa luận.

Khóa luận là kết quả của một quá trình kết hợp nghiên cứu những kỹ thuật trong lĩnh vực tính toán thông minh và nghiệp vụ chứng khoán. Để có thể áp dụng được các kỹ thuật trong tính toán thông minh vào giải quyết bài toán dự đoán chứng khoán đòi hỏi phải qua một số bước sau:

- Tìm hiểu nghiệp vụ.
- Thu thập thông tin có liên quan đến giá cả của các loại cổ phiếu được niêm yết cho đến hiện tại từ năm 2000 đến hiện tại.
- Chọn lựa những tiêu chí và các chỉ số trong đánh giá chứng khoán nhằm đạt được hiệu quả cao và bao quát.
- Áp dụng trí tuệ nhân tạo vào giải quyết bài toán.

1.4 Hướng tiếp cận và các đối tượng nghiên cứu

1.4.1. Hướng tiếp cận

Trong đầu tư chứng khoán, có 2 xu hướng khác nhau: (1) Phân tích cơ bản (2) Phân tích kỹ thuật. Khóa luận sẽ tiếp cận theo cả hai hướng tuy nhiên sẽ chú trọng hơn vào xu hướng thứ 2 là “Phân tích kỹ thuật”. Hai hướng sẽ có những kỹ thuật được áp dụng khác nhau:

Phân tích cơ bản

➤ *Phương pháp CBR-Kết hợp Logic mờ và mạng nơron*

Phân tích kỹ thuật

➤ *Áp dụng mô hình mạng nơron:*

- *Dự đoán giá sử dụng mô hình ARIMA với giá đóng cửa.*
- *Dự đoán giá sử dụng các chỉ số kỹ thuật.*

➤ *Phương pháp ECBR (Ensemble Case Base Reasoning) một cải tiến nhỏ từ phương*

pháp CBR truyền thống.

➤ *Sử dụng Time-series model kết hợp Fibonacci trong lý thuyết sóng Elliott*

Phân tích kỹ thuật được đề cập khá nhiều trong khóa luận và được áp dụng khá nhiều phương pháp. Phân tích cơ bản trong đầu tư chứng khoán cũng cực kì quan trọng, tuy nhiên đây là bài toán khó và thực sự theo hiểu biết thì chưa có một thuật toán hay một ứng dụng nào giải quyết được vấn đề trên, vì nó khá trừu tượng cũng như phụ thuộc nhiều vào cảm giác, kinh nghiệm của nhà đầu tư nhiều hơn, và mang tính chủ quan. Kinh nghiệm đầu tư chính là một tiêu chí quan trọng trong đầu tư và mô hình CBR có lẽ là khá phù hợp cho khi áp dụng cho giải quyết vấn đề này.

Dựa trên phân tích kỹ thuật với khá nhiều phương pháp được áp dụng, tuy mỗi mô hình có những kết quả đạt được khác nhau, nhưng hầu như những kết quả này cũng có những thành quả cho riêng bản thân nó. Có những phương pháp đạt được kết quả khá tốt nhưng ngược lại cũng có phương pháp chưa thật sự hiệu quả. Tuy nhiên, điều quan tâm là nhất ở đây là những kiến thức mới, những hiểu biết mới trong quá trình thực hiện khóa luận.

1.4.2. Đối tượng nghiên cứu

Vì thị trường chứng khoán thực sự là khá phức tạp nếu xem xét tất cả những khía cạnh của thị trường là vượt quá khả năng hiện tại. Vậy nên, ở đây chỉ tập trung xây dựng hỗ trợ dự đoán ở một mảng nhỏ trên thị trường, chính xác hơn là sẽ dựa vào những chỉ số trên bảng điện tử niêm yết trên thị trường để tính toán và đưa ra những dự đoán và đánh giá hỗ trợ đầu tư. Ngoài ra chỉ hỗ trợ cho dự đoán đối với cổ phiếu, còn đối với những loại khác thì ở đây không bàn tới.

Trong khóa luận, có một số đối tượng thường được chú trọng khá nhiều và phân tích kỹ vì nó là nguồn gốc để đạt được hiệu quả cao. Trong hai lĩnh vực “Phân tích cơ bản” và “Phân tích kỹ thuật” được tiếp cận ở trên, mỗi hướng tiếp cận những đối tượng khác nhau:

- *Hướng phân tích cơ bản*: chú trọng nhiều đến các chỉ số tài chính của các công ty đang sở hữu những cổ phiếu được niêm yết trên sàn.
- *Hướng phân tích kỹ thuật*: Dựa vào những chỉ số giá cả trong quá khứ của các cổ phiếu được niêm yết trên sàn sao dịch. Thường sẽ dự đoán giá đóng cửa của một số cổ phiếu ở các phiên giao dịch trong tương lai. Ngoài ra, sẽ nghiên cứu một số những chỉ số kỹ thuật thường sử dụng trong phân tích kỹ thuật vào việc dự đoán giá đóng cửa cổ phiếu trên thị trường.
- Cuối cùng xây dựng công cụ hỗ trợ phân tích kỹ thuật cho phép người sử dụng có những cái nhìn trực quan hơn trong vấn đề tương tác trong dự đoán chứng khoán.

Chương 2. CÁC KIẾN THỨC NỀN TẢNG

2.1 Thị trường chứng khoán

Trước khi bắt đầu đi vào những vấn đề cơ bản chúng ta cần đi vào một số khái niệm cơ bản, tuy nhiên sẽ không đề cập nhiều vào vấn đề này mà sẽ tập trung chủ yếu vào các phương pháp giải quyết vấn đề ở phần sau:

Chứng khoán: là chứng chỉ và bút toán ghi sổ, xác nhận các quyền lợi hợp pháp của người sở hữu đối với tài sản hoặc vốn của tổ chức phát hành.

Thị trường chứng khoán: là một thị trường mà nơi đó người ta mua hoặc bán, chuyển nhượng, trao đổi chứng khoán nhằm mục đích kiếm lời.

Tiếp theo chúng ta đi vào hai vấn đề chính mà sẽ được sử dụng và phân tích nhiều trong khóa luận.

() : Đơn vị 1000 VNĐ (-) : Đơn vị 10 cổ phiếu/chứng chỉ quỹ

P: 1580 - 13/07/2007

Đợt 1: 1010.71 -6.97 -0.68% Tổng GT: 169,879,000,000 Đợt 2: 1010.84 -6.84 -0.67% Tổng GT: 337,314,000,000 Đợt 3: 1015.73 -1.95 -0.19% Tổng GT: 587,431,000,000

CK	Trần	Sàn	Tham chiếu	Đư mua						Đư bán						Khớp lệnh		+/-	Đợt 1		Đợt 2	
				Giá 1	KL 1	Giá 2	KL 2	Giá 3	KL 3	Giá 1	KL 1	Giá 2	KL 2	Giá 3	KL 3	Giá (*)	KL (**)		Giá	KL	Giá	KL
ABT	114	104	109	109	147	108	78	107	50	110	60	111	130	112	486	109	573	0	108	313	108	120
AGF	110	100	105	104	104	103	494	102	110	105	444	106	10	107	415	104	631	-1.0	103	110	103	200
ALT	112	102	107	105	56					107	110	109	50	110	128	107	176	0	107	32	105	84
BBC	57.5	52.5	55	57.5	11,167	56.5	20	56	5							57.5	8,167	+2.5	57.5	1,798	57.5	1,713
BBT	25.2	22.8	24	24.5	468	24.1	209	24	1,715	24.9	150	25	2,212	25.2	305	24.5	5,482	+0.5	23.7	2,680	24.5	1,530
BF1	12.3	11.3	11.8	11.7	1,115	11.6	1,472	11.5	1,871	11.8	4,045	11.9	181	12	3,080	11.8	8,946	0	11.8	3,131	11.8	2,494
BHS	43	39	41.8	41.5	593	41.3	100	41.1	100	41.6	10	41.7	100	41.8	134	41.5	1,811	-0.3	41	304	41.5	685
BMC	457	415	436	441	516	440	457	439	517	445	125	450	230	455	6	440	2,209	+4.0	437	665	438	782
BMP	201	183	192	193	10	192	93	190	230	195	632	196	874	197	280	195	279	+3.0	192	94	192	126
BPC	36.7	33.3	35	35	180	34.5	100	34.4	1,000	35.7	93	35.8	501	35.9	300	35.8	3,280	+0.8	35	20	35.8	2,000
BT6	68	62	65	64	767	63.5	900	63	1,200	64.5	15	65	954	65.5	1,094	64.5	18,954	-0.5	65.5	9,566	65	6,330
BTC	30.9	28.1	29.5	29.7	1	29.5	170	29	860	29.9	72	30	248	30.5	100	30	907	+0.5	30	379	30	186
CAN	29.7	26.9	28.3	28.7	97	28.6	40	28.5	400	28.9	200	29	500	29.3	30	28.7	853	+0.4	28.3	200	28.3	200
CII	70	64	67	67.5	70	67	580	66	240	68	449	68.5	180	69	993	68	5,972	+1.0	67	1,115	67	2,656
CLC	50.5	45.7	48.1	47.8	16	46.6	20	46.5	50	48	32					48	234	-0.1			47.8	34
COM	72	66	69	68.5	594	68	610	67.5	250	69	100	70	50	71	20	68	521	-1.0	67.5	100	68	116
CYC	17.3	15.7	16.5	16.4	70	16.3	847	16.2	200	16.5	741	16.7	30	16.8	160	16.4	977	-0.1	16.4	100	16.5	577
DCT	32.5	29.5	31	31	493	30.8	280	30.7	460	31.4	315	31.5	595	31.7	67	31	2,213	0	31.4	310	31	554
DHA	84	76	80	81	611	80.5	100	80	253	81.5	620	82	315	82.5	210	81	1,748	+1.0	80	100	80	409
DHG	467	423	445	428	69	427	95	425	2	439	100	440	17	444	456	428	1,018	-17.0	440	180	440	223
DIC	38.9	35.3	37.1	36.7	20	36.6	20	36.5	155	36.8	105	37	270	38	70	36.8	545	-0.3	37	160	36.5	185
DMC	144	132	138	139	965	138	1,419	137	10	140	1,969	141	225	142	1,145	139	9,762	+1.0	139	5,608	138	2,462
DNP	75	68	71.5	73.5	82	73	100	72.5	465	74.5	150	75	60			73.5	1,740	+2.0	71.5	305	72	880
DPC	35.1	31.9	33.5	33.5	220	32.8	20	32	140	34	446	35	45			34	950	+0.5	33.5	300	34	20
DRC	131	119	125	125	405	124	494	123	152	126	170	127	150	128	810	125	1,698	0	125	801	124	350
DTT	57	52	54.5	55	23	54.5	110	54	1,290	56	1,490	56.5	11	57	700	56	3,557	+1.5	54	927	53.5	1,140
DYP	47.2	42.8	45	45	35	44.7	90	44.6	20	45.5	250	45.8	200	45.9	10	44.5	285	-0.5	45.5	210	44.5	30

Hình 1 : Một phần bảng giao dịch trực tuyến TP HCM 13/07/2007
(Nguồn www.vnexpress.net)

2.1.1. Phân tích cơ bản

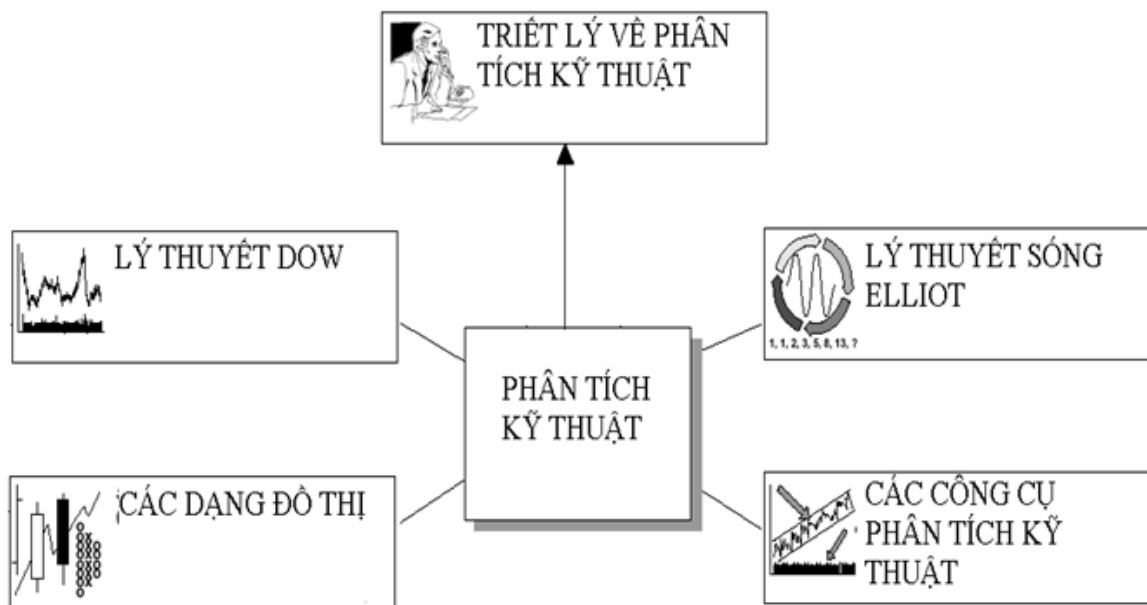
Là việc phân tích bản cân đối tài khoản và bản báo cáo lợi tức của công ty để xem xét chất lượng của công ty cũng như việc phát triển của công ty theo thời gian, nhờ đó tiên đoán các chuyển biến giá chứng khoán. Việc phân tích cơ bản sẽ đánh giá một chứng khoán dưới giá trị hay trên giá trị hiện hành. Có 4 bước chính sau [13],[15]:

- Phân tích điều kiện nền kinh tế nói chung
- Xác định điều kiện ngành công nghiệp
- Xác định điều kiện công ty
- Xác định giá trị cổ phiếu

2.1.2. Phân tích kỹ thuật

Là nghiên cứu số cung và cầu chứng khoán dựa trên các nghiên cứu số lượng và giá cả. Nhà phân tích dùng các biểu đồ để phân tích chiều hướng giá. Hầu hết các phân tích được thực hiện trong thời gian ngắn và trung hạn. Không giống như phân tích cơ bản, nhà phân tích kỹ thuật không quan tâm đến vị thế tài chính của công ty.

Để nghiên cứu về phân tích kỹ thuật và có thể áp dụng hiệu quả cho việc đầu tư chứng khoán phải có những kiến thức nền tảng sau. Xem hình 2 dưới đây:



Hình 2 : Các kiến thức liên quan đến phân tích kỹ thuật

Để có đầy đủ những hiểu biết trên cần một khoảng thời gian để tìm hiểu, nghiên cứu. Ngoài ra trong khóa luận chỉ chủ yếu đề cập vào lĩnh vực tin học nhằm phát triển các phương pháp dự đoán hơn và những vấn đề lý thuyết kinh tế, nên những vấn đề này sẽ không được đề cập chi tiết. Có thể tham khảo [1],[5],[6],[8],[11],[14],[15],[20] để có những kiến thức sâu hơn trong lĩnh vực này.



Hình 3 : Một biểu đồ thể hiện phân tích kĩ thuật chỉ số VNINDEX

2.2 Suy diễn dựa vào tình huống - CBR (Case-Based Reasoning)

2.2.1. Giới thiệu

Trước khi đi vào tìm hiểu chi tiết các kĩ thuật trong CBR ta sẽ tìm hiểu một số các định nghĩa cơ bản. Có thể xem thêm trong [16],[19],[46] để có những tìm hiểu sâu hơn.

Suy diễn dựa vào tình huống: là suy diễn từ các tình huống có sẵn. Tình huống có sẵn ở đây là các tình tiết, các trường hợp hay các kinh nghiệm trong quá khứ. Hệ suy diễn dựa tình huống sẽ dùng những tình huống này để đưa ra giải pháp cho tình huống mới. Giải pháp có thể là một lời giải hoàn chỉnh, một phương pháp thích nghi, một lời cảnh báo, một sự phân loại tình huống,...

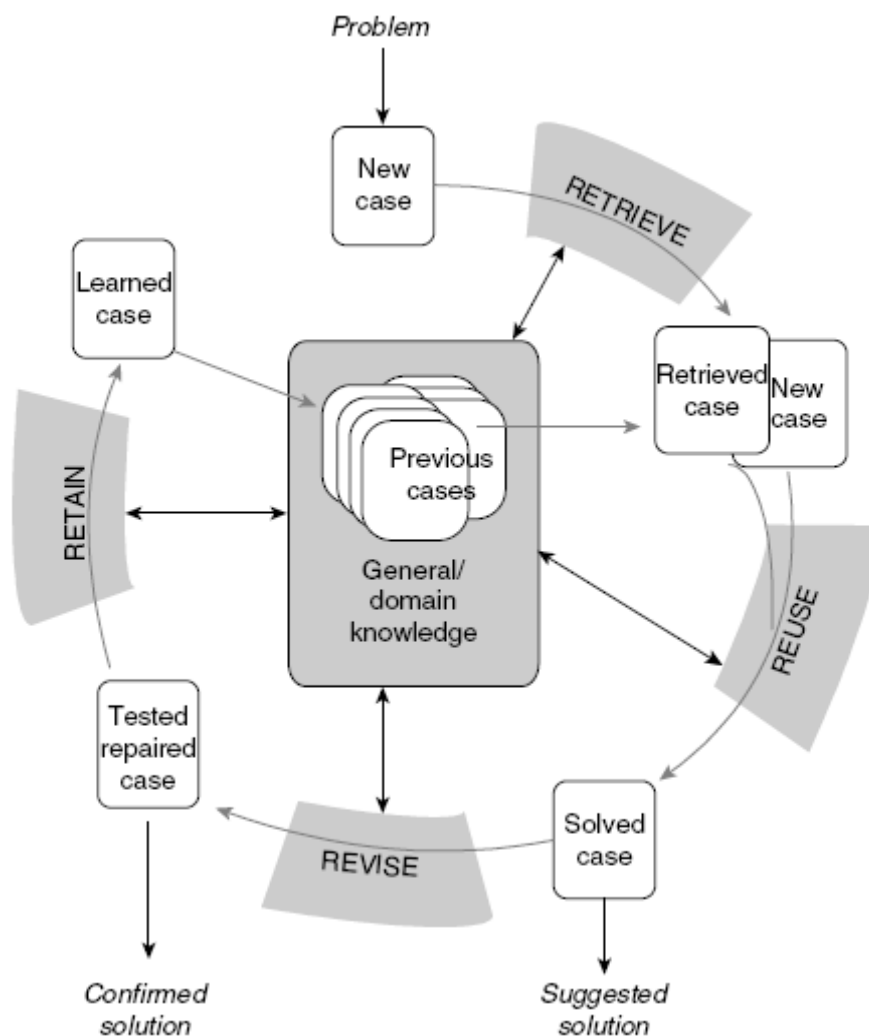
Phương pháp suy diễn dựa tình huống dựa vào theo [46] có 3 giả định sau :

- Các hành động tiến hành trong điều kiện giống nhau hay tương tự nhau thường dẫn đến các kết quả giống nhau hay tương tự nhau.

- Các sự kiện có xu hướng lặp lại. Như vậy các kinh nghiệm trong hệ CBR thường hữu ích trong tương lai.
- Những thay đổi nhỏ trong thế giới chỉ yêu cầu những thay đổi nhỏ trong nhận thức của chúng ta về thế giới, và những thay đổi nhỏ trong cách mà chúng ta thích nghi những tình tiết thay đổi này.

2.2.2. Các thành phần của CBR

Cấu trúc của một CBR cơ bản gồm 4 tác vụ chính sau [19]:



Hình 4 : Chu kì CBR

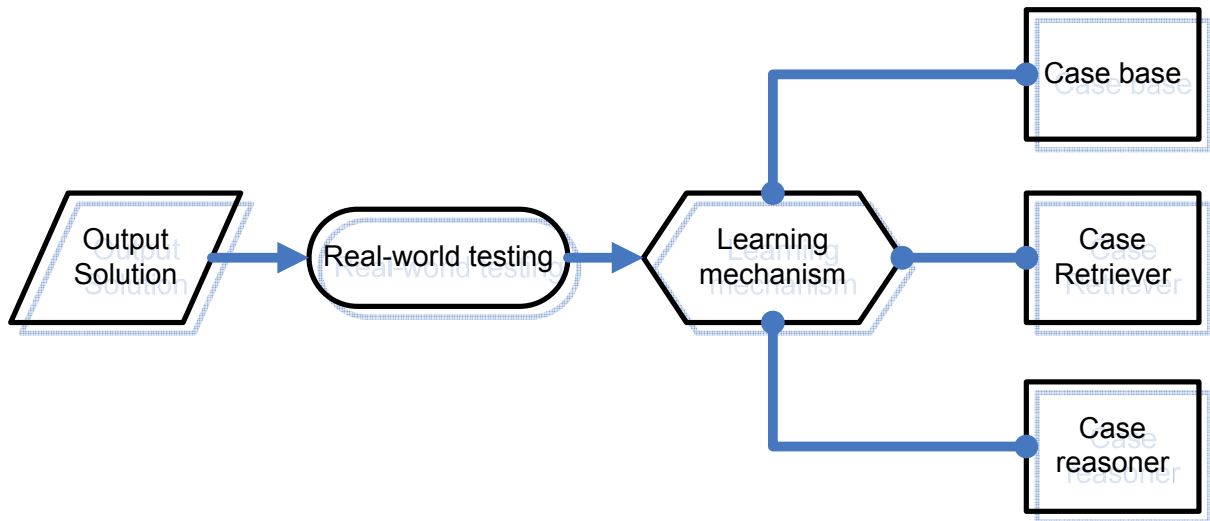
- Rút trích (retrieve) độ “tương tự” của những tình huống trước đó trong quá khứ mà khá giống với vấn đề mà chúng ta cần giải quyết.
- Tái sử dụng (reuse) các tình huống bằng cách sao chép hoặc tổng hợp những lời giải từ các tình huống được rút trích.
- Duyệt lại (revise) hoặc thích nghi (adapt) những lời giải được rút trích nhằm cố gắng để giải quyết vấn đề mới.
- Lưu (retain) vấn đề mới xuống khi nó được xác nhận và kiểm tra hợp lệ.

Trong một số ứng dụng thực tế thì, hai giai đoạn “Tái sử dụng” và “Duyệt lại” thường là khó phân biệt, và một số nhà nghiên cứu sử dụng giai đoạn thích nghi nhằm thay thế cho hai giai đoạn kia. Tuy nhiên, giai đoạn thích nghi trong hệ thống CBR vẫn là một câu hỏi mở vì nó là quá trình phức tạp mà cố gắng tính toán lời giải của tình huống. Vấn đề này yêu cầu sự phát triển mô hình quan hệ nhân quả giữa không gian vấn đề và không gian lời giải liên quan đến tình huống.

Trong chu kì của CBR ở trên ta thấy, các tình huống được lưu trữ trong các cơ sở tình huống mà được bổ sung bởi những tri thức chung, mà thường là độc lập miền.

Cách nhìn hướng xử lí của CBR, của chu kì CBR cung cấp một khả năng xem xét cái gì đang xảy ra, trong khi cái nhìn hướng tác vụ của CBR thì tốt cho việc miêu tả những kĩ thuật thật sự. Các tác vụ thường dựa trên mục đích của hệ thống, và một số tác vụ đặc biệt được thực hiện bằng cách áp dụng một hoặc nhiều phương thức.

Các tác vụ của một CBR rất phức tạp nhưng có thể đơn giản hóa bằng lược đồ sau:



Hình 5 : Các thành phần cơ bản của CBR

2.2.2.1. Biểu diễn, đánh chỉ số tình huống (Case representation and indexing)

Như đã đề cập ở trước, một tình huống có thể được lưu như là một kinh nghiệm hay một vấn đề trước đó, nó sẽ được lưu phụ thuộc vào miền cũng như phụ thuộc vào mục đích mà tình huống sẽ được sử dụng. Đối với hệ thống giải quyết vấn đề, chi tiết sẽ bao gồm đặc tả của vấn đề các thuộc tính liên quan miêu tả ngữ cảnh của vấn đề. Một phần quan trọng khác của một tình huống là việc miêu tả lời giải được sử dụng trong mối quan hệ trước đây khi gặp phải một lời giải tương tự. Phụ thuộc vào hệ thống CBR quyết định với các tình huống, lời giải này có lẽ bao gồm chỉ những sự kiện mà định nghĩa lời giải, hoặc nó có thể bao gồm thông tin về các bước hoặc xử lý thêm liên quan đến việc đạt được lời giải. Nó cũng quan trọng khi bao gồm độ đo sự thành công của việc miêu tả tình huống nếu một lời giải hay một tình huống trong cơ sở trường hợp đạt được ở cả mức độ thành công hay thất bại.

Khi so sánh giữa việc lưu trữ tri thức trong một hệ thống dựa vào luật và hệ thống lưu trữ dạng tình huống. Trong khi tri thức trong mô hình dựa vào luật có tính trừu tượng để nó có thể áp dụng trong nhiều trường hợp thì tri thức được lưu trữ trong hệ thống dựa vào cơ sở tình huống sẽ giữ lại những thuộc tính đặc biệt của một tình

huống ở nơi mà nó lưu trữ, nghĩa là nó là những tri thức liên quan chỉ áp dụng trong một ngữ cảnh xác định nào đó. Do đó thay vì phải đưa ra những hệ tri thức rộng, phức tạp thì ta có thể xác định những tri thức nhằm giải quyết những vấn đề đặc biệt có thể nhóm chúng lại trong một hoặc một vài tình huống.

Cơ sở tình huống trong hệ thống CBR là bộ nhớ của tất cả các tình huống được lưu trữ trước đây. Có 3 vấn đề chung mà phải xác định nên tạo một cơ sở tình huống khi nào [19]:

- Cấu trúc biểu diễn tình huống.
- Mô hình bộ nhớ được sử dụng để tổ chức toàn bộ cơ sở tình huống.
- Lựa chọn phương pháp đánh chỉ số để có thể xác định mỗi tình huống.

Biểu diễn tình huống (Case representation)

Trong mỗi hệ CBR việc lưu trữ cơ sở tình huống vẫn được nghiên cứu sao cho việc lưu trữ đạt được hiệu quả, nó là vấn đề không phải là đơn giản. Với các phương thức truyền thống, chúng ta có 3 phương pháp sau:

- Biểu diễn theo quan hệ (Rational Representation).
- Biểu diễn theo hướng đối tượng (Object-oriented Representation).
- Biểu diễn theo tính chất (Predicate Representation)

Đánh chỉ số tình huống (Case indexing)

Trong hầu hết các cơ sở dữ liệu lưu trữ đều phải được đánh chỉ số nhằm tăng tốc độ rút trích dữ liệu. CBR cũng vậy, các cơ sở tình huống cũng được đánh chỉ số với cùng mục đích này:

- Thông tin được đánh chỉ số được sử dụng trong quá trình rút trích.
- Thông tin không được đánh chỉ số có nhiệm vụ cung cấp các thông tin ngữ cảnh có giá trị đến người dùng nhưng không được sử dụng trực tiếp trong quá trình rút trích.

Ngoài ra việc đánh chỉ số cũng giúp ta lựa chọn những tình huống một cách đúng đắn. Điều này dẫn tới các thuộc tính được đánh chỉ số phải thể hiện là các thuộc tính quan trọng của tình huống, nghĩa là các chỉ số phải phản ánh giá trị đầu ra của một tình huống và miêu tả ngữ cảnh mà một tình huống có thể rút trích trong tương lai. Ngoài ra thuộc tính được đánh chỉ số phải đủ trừu tượng nhưng cũng đừng quá trừu tượng [19].

2.2.2.2. Rút trích tình huống (Case Retrieval)

Đây là quá trình tìm kiếm tình huống gần nhất với tình huống hiện tại. Để đạt được hiệu quả rút trích chọn lựa tốt nhất, chúng ta cần có một tiêu chuẩn chọn lựa mà xác định một tình huống nào là thích hợp cho việc rút trích và là một bộ phận để điều khiển một cơ sở tình huống được tìm thấy như thế nào. Tiêu chuẩn chọn lựa là cần thiết để xác định tình huống nào là thích hợp để rút trích, bằng cách xác định mức độ gần như thế nào với các tình huống lưu trữ.

Chọn lựa tình huống phụ thuộc một phần vào bộ rút trích tình huống tìm kiếm cái gì trong cơ sở tình huống. Thường thì bộ rút trích tìm kiếm toàn bộ tình huống cũng như các đặc trưng mà được so sánh với tình huống hiện hành.

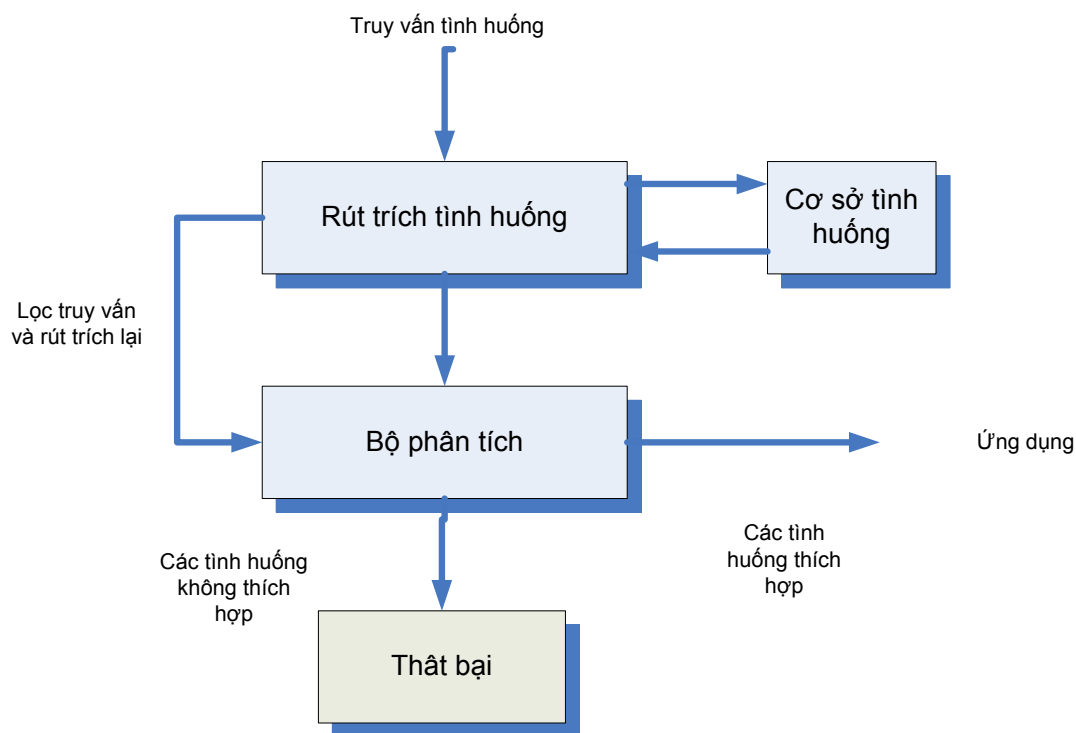
Có một số nhân tố nhằm xem xét khi nào xác định phương pháp rút trích:

- Số tình huống được tìm kiếm.
- Lượng tri thức miền có sẵn.
- Mức độ dễ dàng trong việc xác định trong số cho các thuộc tính.
- Tất cả các tình huống nên được đánh chỉ số trên cùng một số thuộc tính, hoặc xác định xem tình huống những thuộc tính nào quan trọng.

☛ Khi xem xét phương pháp phân tích có một số điểm đáng lưu ý như sau:

- Thời gian và tài nguyên được yêu cầu cho quá trình thích nghi.
- Số tính huống trong một cơ sở tình huống.
- Thời gian và tài nguyên được yêu cầu cho việc tìm kiếm.
- Bao nhiêu cơ sở tình huống được tìm kiếm.

Quá trình này được mô tả bằng sơ đồ sau:

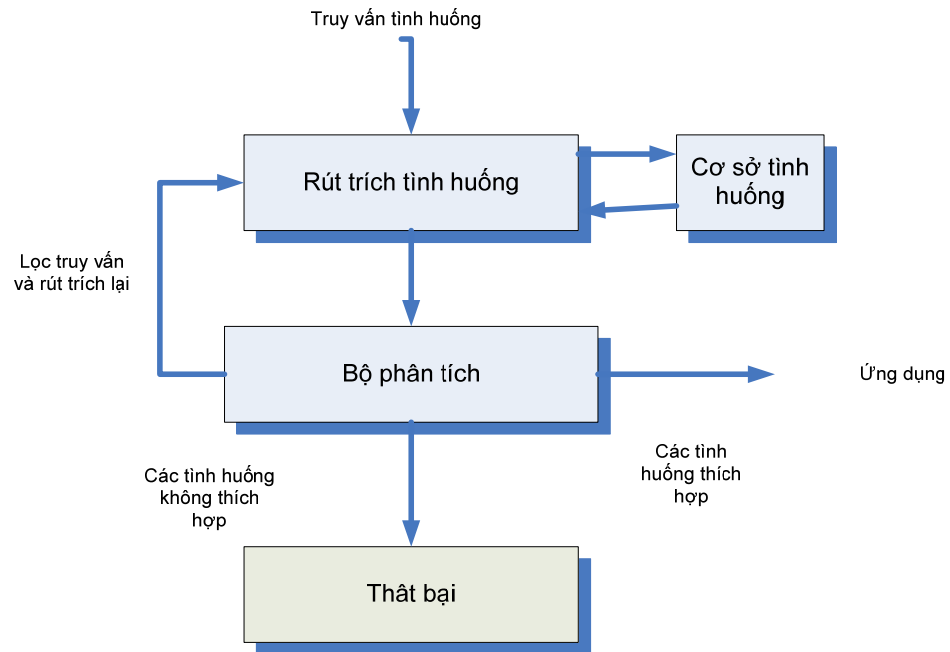


Hình 6 : Quá trình rút trích tình huống

2.2.2.3. Sự thích nghi tình huống (Case adaptation)

Là quá trình chuyển đổi một lời giải được rút trích thành một lời giải thích hợp cho vấn đề hiện tại. Nó được chỉ ra rằng quá trình này có thể là một bước quan trọng nhất của

CBR vì nó thêm những thông tin tri thức thông minh để có thể đơn giản hóa việc đối



sánh các mẫu.

Hình 7 : Quá trình rút trích tình huống

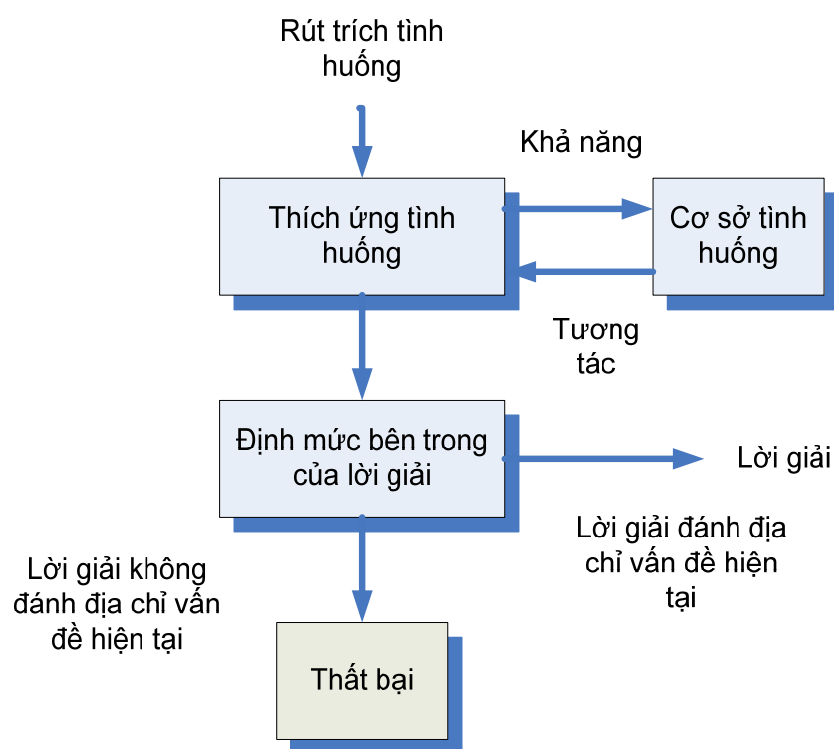
Một số các phương pháp được sử dụng để tiến hành quá trình thích ứng:

- Một lời giải được trả về có thể được sử dụng như là một lời giải cho vấn đề hiện tại mà không cần phải sửa đổi, hoặc có thể sửa đổi mà ở đó lời giải không hoàn toàn tương thích với vấn đề hiện hành.
- Các bước hoặc các quá trình được cho phép để đạt được một lời giải sớm có thể được chạy lại mà không cần phải sửa đổi hoặc có sửa đổi nhưng các bước đạt được trong tình huống trước đó là không hoàn toàn thỏa mãn với tình huống hiện hành.
- Nếu có nhiều hơn một tình huống cho rút trích, một lời giải có thể được đưa ra từ nhiều tình huống hoặc nói cách khác, một vài các lời giải khác cũng có thể được biểu diễn.

Quá trình này có sử dụng nhiều kỹ thuật khác nhau, bao gồm các luật hoặc việc ra quyết định xa hơn. Khi chọn các chiến lược cho quá trình này, nên theo một trong những phương pháp sau:

- Dựa vào khoảng cách tình huống được rút trích so với tình huống hiện hành gần nhau như thế nào? Các phương pháp tính khoảng cách có thể sử dụng như: Manhattan, Euclide v.v...
- Một cách đơn giản hơn, tính xem có bao nhiêu thuộc tính khác nhau giữa các tình huống.
- Có một sự chung hay một số các luật được biết khác có thể được sử dụng khi tiến hành quá trình này.

Sơ đồ quá trình này như sau:



Hình 8 : Quá trình thích ứng tình huống

2.2.2.4. Bảo trì tình huống

Bảo trì cơ sở tình huống liên quan đến các kỹ thuật thêm, bớt, cập nhật các tình huống, các chỉ số, và các cấu trúc tri thức khác trong hệ thống CBR nhằm đảm bảo tính đúng đắn và khả năng thi hành của hệ thống. Các tác vụ bảo trì của CBR có thể chia làm 2 loại: (1) bảo trì định tính, (2) bảo trì định lượng.

Bảo trì định tính: nhằm đảm bảo tính hiệu quả của hệ thống bao gồm: tính đúng đắn, tính vững vàng, và hoàn hảo của hệ thống.

Bảo trì định lượng: hướng tới hiệu quả giải quyết vấn đề của hệ thống CBR như: Cải tiến thời gian giải quyết trung bình của hệ thống, điều khiển kích thước của cơ sở tri thức, các tổ chức cấu trúc chỉ số.

Có khá nhiều kỹ thuật được áp dụng nhằm thực hiện với những bảo trì đã nêu ở trên như sử dụng GA, Logic mờ, mạng nơron v.v... Ở đây chỉ trình bày vắn tắt những vấn đề cơ bản như là một giới thiệu có thể tham khảo [16] để có những tìm hiểu chi tiết hơn về những phương pháp này.

Chương 3. CÁC PHƯƠNG PHÁP VÀ MÔ HÌNH DỰ ĐOÁN

3.1 Áp dụng trong phân tích cơ bản

3.1.1. Đánh giá cổ phiếu dựa trên phân tích cơ bản

Phân tích cơ bản, ngay từ đầu đã là một trong những phương pháp dựa vào cảm nhận, kinh nghiệm của mỗi nhà đầu tư để ra quyết định. Những nhà đầu tư sành sỏi thường dựa trên những kiến thức cũng như những kinh nghiệm đã thu nhặt được trong những khoảng thời gian đầu tư trước đó. Những kinh nghiệm này hình thành những qui tắc, những luật trong đầu tư của họ. Một nhà đầu tư tài chính nói chung hay một nhà đầu tư chứng khoán dựa vào những phân tích cơ bản nói riêng thường dựa vào những tình huống trong quá khứ đã biết hay nói đúng hơn là những kinh nghiệm quá khứ để ra quyết định. Ví dụ, như việc đánh giá giá trị của chứng khoán, mức độ tốt của chứng khoán. Thường thì phân tích cơ bản được sử dụng trước trong đầu tư chứng khoán, sau đó người ta sẽ dựa vào những phân tích kỹ thuật áp dụng trên từng loại cổ phiếu để ra những quyết định của mình.

Thị trường chứng khoán Việt Nam là một thị trường hoạt động đến nay chỉ khoảng 7 năm và có nhiều biến chuyển. Những sự biến động hỗn loạn của thị trường những năm gần đây, chính bản thân thị trường cũng thể hiện sự non trẻ của nó. Đặc biệt là thời gian gần đây thị trường cực kì biến động và mất ổn định và phải có sự tác động của nhà nước để ổn định thị trường, sự phát triển mạnh của thị trường không đồng nghĩa với sự ổn định tương ứng của nó. Với những người chưa có kinh nghiệm và chạy theo phong trào thì sẽ rất khó khăn, ngay từ đầu khi xây dựng khóa luận đã chú trọng tới vấn đề này. Các chuyên gia thực sự am hiểu thị trường chứng khoán Việt Nam thực sự không nhiều, vì vậy khó để có thể đưa ra các luật hiệu quả, các quyết định hiệu quả trong chiến lược đầu tư.

Dựa trên những phân tích ở trên trong, phần này sẽ mô hình hóa chiến lược bằng hệ quyết định CBR. Lý do tại sao sử dụng mô hình này mà không sử dụng mô hình khác đã được trình bày ở trên. Dưới đây là phân mô tả chi tiết và phương pháp xây dựng mô hình ứng dụng dựa trên CBR.

3.1.1.1.1. Chức năng và định hướng của phương pháp

Mục tiêu chính của phần phân tích cơ bản này là xây dựng một mô hình đánh giá cơ bản giá trị của một cổ phiếu, và triển vọng của nó trong tương lai. Tức là bước đầu trong việc đánh giá cổ phiếu tốt hay xấu, có triển vọng trong tương lai hay không. Mục tiêu của phương pháp này không thực sự giúp người sử dụng biết nên mua bao nhiêu loại cổ phiếu, hay nên mua bao nhiêu cổ phiếu trong một loại, nên đầu tư bao nhiêu tiền để mua cổ phiếu... Mà như đã nói ở trên mục tiêu này chỉ bao gồm việc đánh giá một cổ phiếu có những triển vọng trong tương lai sắp tới hay không. Đó chính là những bước tiếp cận đầu tiên trong việc giải quyết vấn đề.

3.1.1.1.2. Biểu diễn tình huống

Phân biểu diễn tình huống bao gồm 12 thuộc tính và lời giải của bài toán là một trong 3 chọn lựa. Tốt, Trung bình hay Xấu (hay có thể hiểu Không triển vọng – Bình thường – Triển vọng)? Khi được hỏi về chức năng của hệ thống. Một số người cho rằng khi mua chứng khoán phải xảy ra 4 trường hợp sau: Mua – Không mua – Bán ra – Giữ lại vì vậy nếu kết quả chỉ xảy ra 3 trường hợp thì liệu có giúp nhà đầu tư có nên mua hay không nên mua cổ phiếu. Nhưng ở đây có một vấn đề xin được trình bày ở đây như sau:

Một nhà đầu tư chứng khoán khi muốn đầu tư vào một loại cổ phiếu sẽ có 2 trường hợp xảy ra: Người đó chưa mua chứng khoán này và muốn xác định mình có nên mua hay không? Nếu người đó đã có chứng khoán trong tay thì người đó xác định rằng có nên bán ra hay giữ lại để bán với giá cao hơn sau này. Vì vậy ta chỉ cần một lời giải gồm 2 giải pháp Mua – Không mua (trường hợp người chưa có loại cổ phiếu này) hoặc Bán –

Giữ lại (Trường hợp nhà đầu tư đã nắm giữ loại cổ phiếu này). Tuy nhiên, ở đây sẽ không định hướng như trên mà sẽ phân loại cổ phiếu thành 3 kết quả khác nhau chỉ ở mức độ đánh giá cổ phiếu đó thuộc loại nào? Ví dụ: Nếu cổ phiếu đó là loại “Tốt” thì tất nhiên nhà đầu tư nên mua loại cổ phiếu này vì trong tương lai nó có nhiều triển vọng, và ngược lại nếu cổ phiếu “xấu” thì tất nhiên là có thể không nên mua. Ở đây chỉ là một ví dụ nhỏ, có thể còn tùy vào mỗi cá nhân mà sẽ có những chiến lược khác nhau tùy vào mỗi thời điểm. Trong phần này sẽ không giúp người sử dụng có nên mua cổ phiếu hay không? Mà như đã trình bày chỉ đánh giá loại cổ phiếu đó có tốt hay không mà thôi!

Case {

Pre:

- **Chỉ số EPS**
- **Chỉ số P/E**
- **Chỉ số P/B**
- **Chỉ số ROE.**
- **So với chỉ số P/E ngành.**
- **So với chỉ số P/E ngành.**
- **Lợi nhuận công ty từ đâu năm.**
- **Lợi nhuận cung kì trong quá khứ**
- **Khả năng cạnh tranh.**
- **Ưu đãi.**
- **Xu hướng tăng giá**
- **Tính thành khoản.**

Post:

- **Loại cổ phiếu (Tốt – TB – Xấu).**

}

Bảng 1 : Biểu diễn tình huống trong phân tích đánh giá chứng khoán

< **Chỉ số EPS (Earning per share) – Chỉ số thu nhập trên đơn vị cổ phiếu:** thu nhập dòng hay lợi nhuận sau thuế có thể được xem xét bằng tổng lợi nhuận hay xét trên một đơn vị cổ phiếu gọi là chỉ số EPS. Một công ty có sự gia tăng về tốc độ tăng trưởng thu nhập khi tốc độ tăng trưởng thu nhập của quý sau cao hơn quý trước. Một cổ phiếu tốt, có tốc độ gia tăng về tăng trưởng thu nhập cao hơn so với ba hay bốn quý trước. Tăng trưởng của chỉ số thu nhập ít nhất là 25% so với cùng quý của năm trước. EPS hàng năm của cổ phiếu tốt ít nhất phải tăng trưởng 25% so với ba năm trước.

< **Chỉ số P/E (Price/Earnings Ratio) – Hệ số giá và thu nhập cổ phiếu:** P/E là một tỷ lệ phần trăm giữa thị giá một cổ phiếu và thu nhập trên một cổ phiếu. Tỷ số này được tính bằng cách lấy giá giao dịch cổ phiếu chia cho doanh lợi trên mỗi đầu cổ phiếu. P/E có lẽ là một trong những chỉ số quan trọng nhất trong phân tích căn bản. Đối với một số nhà đầu tư thì đây là loại tỷ số hữu ích nhất. Sở dĩ phổ biến như vậy bởi vì P/E là chỉ tiêu dự báo cấp thời đáng tin cậy. P/E cho nhà đầu tư biết rằng nhà đầu tư phải trả bao nhiêu tiền để nhận được 1 đồng trong khoản tiền lãi ấy. Chỉ tiêu này cũng dự báo cho thấy cổ phiếu là đối tượng khả quan hay thấp kém? Thông thường, sau khi tính được tỷ số P/E nhà đầu tư sẽ so sánh với 4 chuẩn mực sau đây:

- Tỷ số tăng trưởng trong quá khứ của loại cổ phiếu ấy.
- Mức tăng trưởng dự kiến tương lai của công ty.
- Cổ phiếu của các công ty khác trong cùng một ngành kinh doanh.
- Toàn bộ thị trường phản ánh các loại chỉ số.
- Tiềm năng phát triển của công ty.

Dựa trên những so sánh này thì nhà đầu tư có thể biết được giá trị thực nhỏ hơn hay lớn hơn sự đánh giá của thị trường.

< **Chỉ số P/B (Price to Book ratio) - Tỷ số giá/Giá cổ phiếu:** Đây là tỷ lệ được sử dụng để so sánh giá của một cổ phiếu so với giá trị ghi sổ của cổ phiếu đó. Tỷ lệ này được tính toán bằng cách lấy giá đóng cửa hiện tại của cổ phiếu chia cho giá trị ghi sổ tại quý gần nhất của cổ phiếu đó.

Đây là tỉ lệ được sử dụng để so sánh giá của một cổ phiếu so với giá trị ghi sổ của cổ phiếu đó. Tỉ lệ này được tính toán bằng cách lấy giá đóng cửa hiện tại của cổ phiếu chia cho giá trị ghi sổ tại quý gần nhất của cổ phiếu đó.

< **Chỉ số ROE (Return on equity - Chỉ số quay vòng vốn):** là chỉ số đo lường mức lợi nhuận đạt được trên đồng vốn đóng góp của các cổ đông, được tính bằng cách lấy lợi nhuận ròng của doanh nghiệp chia cho số vốn cổ đông: $ROE = \text{Lợi nhuận trên tổng tài sản} / \text{Vốn chủ sở hữu}$. Hiện nay các nhà đầu tư được sử dụng khá nhiều, và thường thì khuyến khích đầu tư vào những cổ phiếu có ROE cao hoặc tăng lên. Về mặt lý thuyết những cổ phiếu có ROE càng cao thì thể hiện công ty sử dụng vốn liếng của bạn để đầu tư có hiệu quả. Các cổ phiếu có ROE cao rất cực sẽ được các nhà đầu tư ưa chuộng.

< **So với P/E ngành:** là so chỉ số P/E của cổ phiếu với chỉ số P/E chung của ngành mà công ty phát hành loại cổ phiếu này hoạt động.

< **So với P/B ngành:** tương tự là so chỉ số P/E của cổ phiếu với chỉ số P/E chung của ngành mà công ty phát hành loại cổ phiếu này hoạt động.

< **Lợi nhuận công ty từ đầu năm:** Đánh giá tỉ lệ chuyển doanh thu thành thu nhập, Theo quan điểm của nhà đầu tư thì nên tìm những công ty có sự tăng về lợi nhuận ròng/doanh thu. Con số này càng lớn thì càng thể hiện công ty đang có xu hướng quản lý tốt và các hoạt động khác tốt lên. Tiêu chí để xác định chỉ tiêu này là lợi nhuận sau thuế ít nhất là 10%.

< **Lợi nhuận cùng kì trong quá khứ:** Cũng tính bằng phương pháp trên nhưng là lợi nhuận cùng kì tương đương ở năm trước.

< **Khả năng cạnh tranh:** khả năng cạnh tranh này là so với những công ty cùng ngành cùng lĩnh vực trên thị trường.

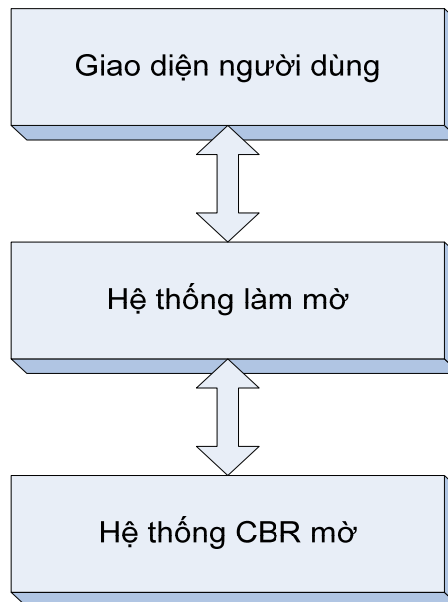
< **Xu hướng tăng giá:** là khả năng mà trong khoảng thời gian đầu tư mà nhà đầu tư xác định trước đó, có khoảng tăng giá so với giá hiện hành.

< **Mức độ ưu đãi:** là mức độ ưu đãi của nhà nước đối với lĩnh vực hoạt động của công ty.

< **Tính thanh khoản:** Nhờ có thị trường chứng khoán các nhà đầu tư có thể chuyển đổi các chứng khoán họ sở hữu thành tiền mặt hoặc các loại chứng khoán khác khi họ muốn. Khả năng thanh khoản là một trong những đặc tính hấp dẫn của chứng khoán đối với người đầu tư. Đây là yếu tố cho thấy tính linh hoạt, an toàn của vốn đầu tư. TTCK (Thị trường chứng khoán) hoạt động càng năng động và có hiệu quả thì tính thanh khoản của các chứng khoán giao dịch trên thị trường càng cao.

Những chỉ số này ngày nay được sử dụng khá phổ biến nhằm đánh giá giá trị của một cổ phiếu, ngay cả các nhà đầu tư Việt Nam cũng rất chuộng những chỉ số này. Trong biểu diễn tình huống có một vấn đề cần chú ý là mỗi tình huống sẽ được gán một trọng số nhất định, trọng số này sẽ do người sử dụng quyết định. Ngoài ra, phương pháp biểu diễn tình huống ở trên là “thô”. Sau khi biểu diễn nó sẽ được làm mờ với các mục đích khác nhau.

3.1.1.1.3. Hệ thống làm mờ (Fuzzifier)

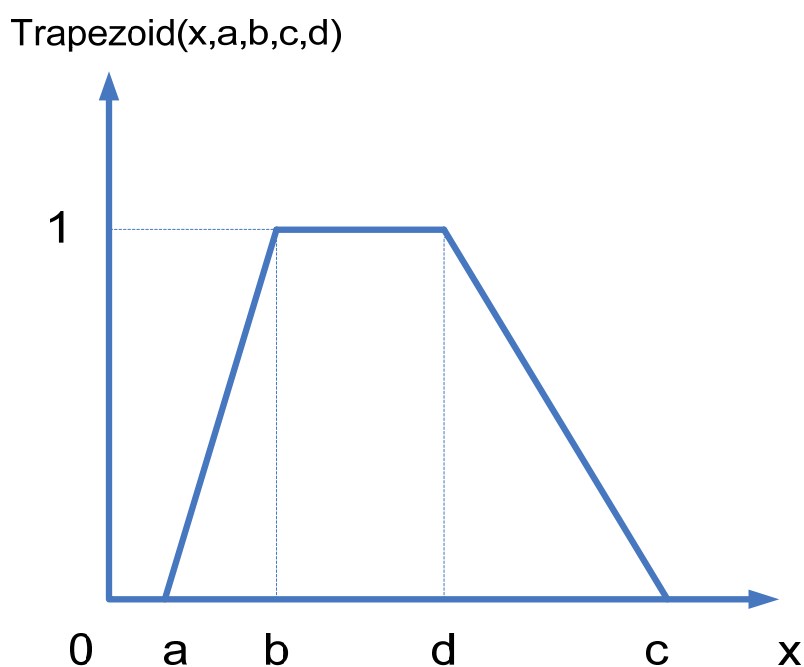


Hình 9 : Hệ thống mờ

Trước khi đi vào thực hiện đưa tình huống vào hệ thống chúng ta phải thực hiện làm mờ. Phương pháp làm mờ là phương pháp hiệu quả nhằm mô tả những biến ngôn ngữ. Trong hệ thống chứng khoán thì việc miêu tả một biến ngôn ngữ sẽ có lợi hơn nhiều trong việc miêu tả bằng con số thô cụ thể. Ví dụ: Giả sử với chỉ số rủi ro là 5% thì thay thì người ta không quan tâm nhiều về giá trị rõ ràng, mà thường người ta đánh giá cổ phiếu đó có tỉ lệ rủi ro cao, thấp hay bình thường.

Phần này sẽ trình bày một cách chi tiết phương pháp mờ hóa các giá trị mờ của từng thuộc tính. Hàm mờ được sử dụng ở đây là hàm trapezoid:

$$\text{trapezoid}(x; a, b, c, d) = \max \left\{ \min \left\{ \frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right\}, 0 \right\}$$



Hình 10 : Hàm mờ trapezoid

Dưới đây là bảng các phân chia thuộc tính mờ:

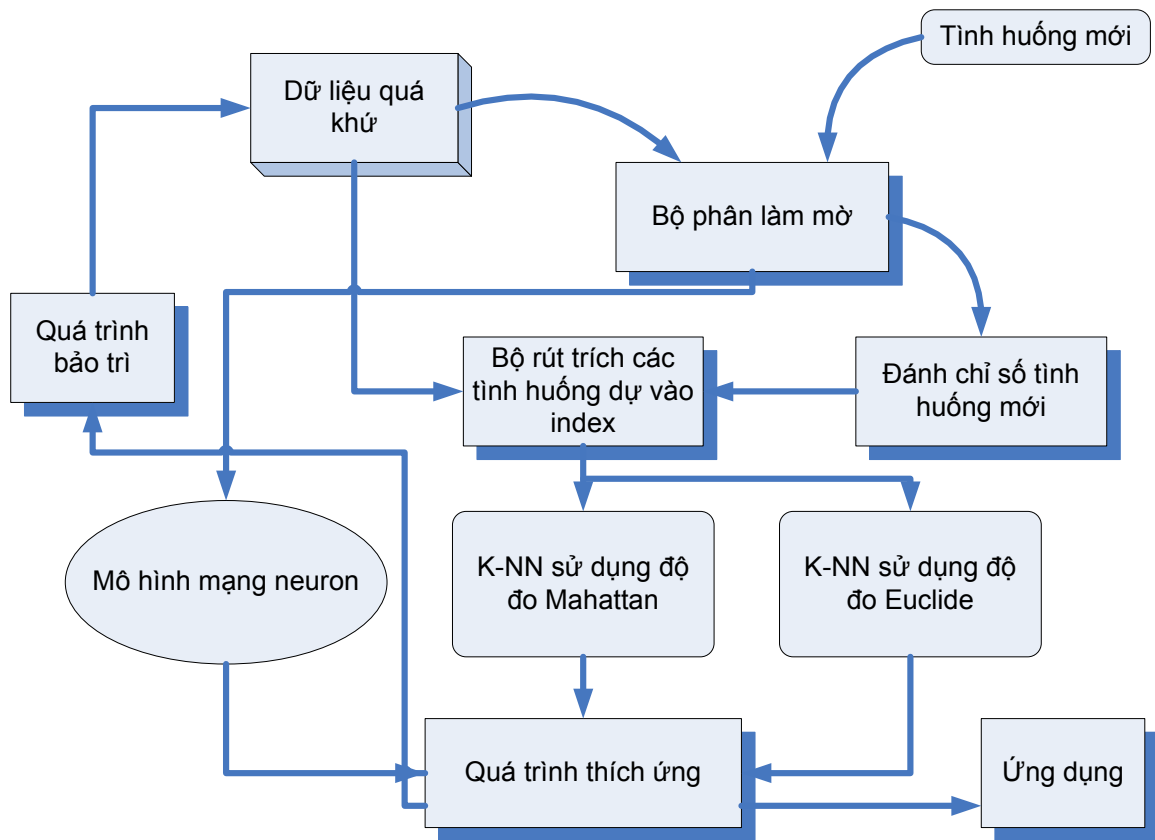
Các thuộc	Miền giá trị	Giá trị biến mờ	Các giá trị ứng với
-----------	--------------	-----------------	---------------------

tính		mỗi giá trị mờ	
EPS	[0 - 100] %	Thấp – TB – Cao	[15,35,100]
P/E	[0 - 500]	Thấp – TB – Cao	[20,40,500]
P/B	[0 - 100]	Thấp – TB – Cao	[2.5,10,100]
ROE	[0 - 100]%	Thấp – TB – Cao	[20,40,100]
So với P/E ngành		Thấp – Tương đương – Cao	
So với P/B ngành		Thấp – Tương đương – Cao	
Lợi nhuận	[0 - 100]%	Thấp – TB – Cao	[9,25,100]
Lợi nhuận cùng kì	[0 - 100]%	Thấp – TB – Cao	[9,25,100]
Cạnh tranh		Yếu – Cạnh tranh – Độc quyền	
Ưu đãi		Có – Không	
Xu hướng giá		Ổn định – Không ổn định	
Thanh khoản		Không – Có	

Bảng 2 : Bảng giá trị mờ hóa các thuộc tính

Không có một tài liệu nào hướng dẫn cách phân chia các khoảng trong hệ thống làm mờ như thế này. Đây chỉ là những kinh nghiệm của những người đã tham gia vào trong thị trường một khoảng thời gian đưa ra những nhận định của họ. Và đây cũng là những bước thử nghiệm đầu tiên khi xây dựng loại bài toán này cho chứng khoán. Có thể sẽ có những phương pháp xác định miền hiệu quả hơn nhưng cũng chưa nghĩ ra phương pháp nào ngoài việc sẽ thực nghiệm những khoảng giá trị này trong thời gian tới.

Mô hình CBR áp dụng cho đánh giá chứng khoán như sau:



Hình 11 : Sơ đồ mô hình CBR cho đánh giá cổ phiếu

Hình 11 là mô hình ứng dụng CBR cho hỗ trợ chứng khoán. Ta sẽ từng bước phân tích đơn giản từng phần trong mô hình để có thể hiểu rõ hơn chi tiết từng bước trong mô hình. Mô hình CBR này tương tự mô hình CBR truyền thống như ở đây có một chút khác biệt:

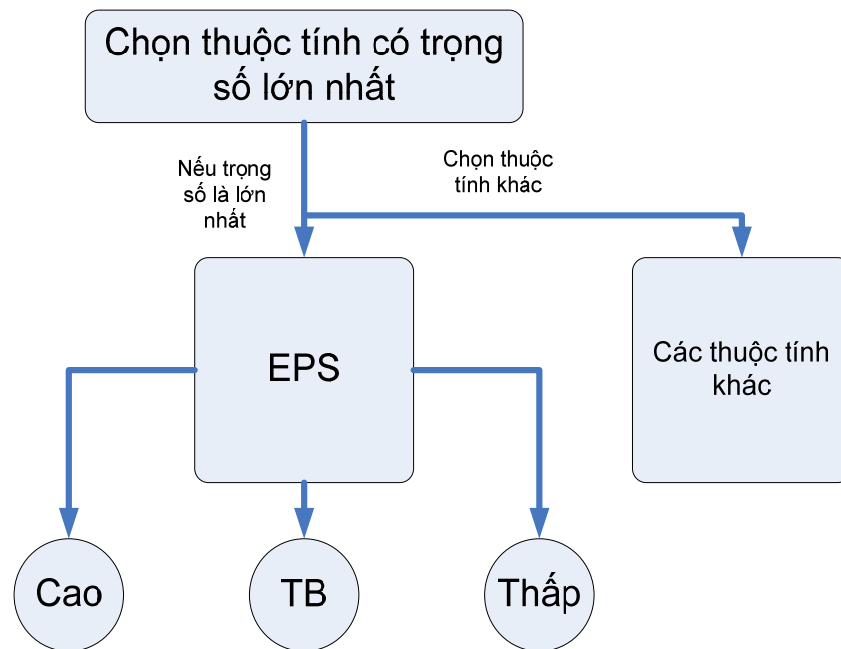
- Thành phần mạng nơron cho phân loại tình huống sẽ thực hiện phân loại một tình huống mới vào thành một trong 3 lớp “Xấu”, “Bình thường”, “Tốt”.
- Thành phần k-NN sử dụng độ đo Manhattan tính khoảng cách gần nhất giữa tình huống hiện hành và các tình huống trong quá khứ.
- Thành phần k-NN sử dụng độ đo Euclide tính khoảng cách gần nhất giữa tình huống hiện hành và các tình huống trong quá khứ, nhưng ở đây có một khác biệt là phương pháp tính này không bị ảnh hưởng bởi các trọng số của thuộc tính.

3.1.1.1.4. Đánh chỉ số tình huống

Trong các hệ CBR truyền thống thì việc đánh chỉ số tình huống dựa trên nhiều mô hình khác nhau, trong những ý tưởng đầu tiên được đề xuất là mô hình hóa quá trình đánh chỉ số tình huống động. Như trong mô hình truyền thống sử dụng phương pháp xây dựng cây quyết định hoặc dựa vào lý thuyết xác suất để mô hình đánh trọng thuộc tính. Đánh chỉ số động tức là mức độ đánh trọng sẽ do người sử dụng định ra, mức độ tin cậy vào mỗi thuộc tính của họ sẽ ảnh hưởng một phần quyết định của hệ thống và ý tưởng như sau: “Mỗi thuộc tính sẽ gắn kèm một trọng số thể hiện mức độ tin cậy của người sử dụng đối với thuộc tính quyết định này”. Dựa vào tính ưu tiên và mức độ đánh trọng sẽ thực hiện việc lựa chọn các ưu tiên rút trích những tình huống nào và như thế sẽ linh động hơn trong quyết định của mình.

Ở đây, việc đánh chỉ số có thể bị trường hợp quá khớp nếu chỉ sử dụng một thuộc tính. Mô hình đánh chỉ số sẽ dựa vào 2 hoặc 3 thuộc tính mà trọng số cao nhất để thực hiện rút trích. Giả sử người sử dụng tin tưởng vào giá trị cổ phiếu ở 2 thuộc tính đó là “Xu hướng tăng giá” và “Tính thành khoản” sẽ đánh trọng 2 giá trị này là cao nhất và khi thực hiện việc rút trích cho k-NN sẽ lấy những tình huống quá khứ mà những thuộc tính tương tự với tình huống hiện tại dựa vào 2 thuộc tính này. Giá trị k được chọn trong bài toán là $k = 5$. Không có một sự tính toán hay chọn lựa giá trị k nào là tốt nhất, ở đây chỉ là đề xuất giá trị k như vậy. Để có thể tính toán được giá trị k nào là tối ưu thì cần có những kiểm nghiệm thực tế và lâu dài.

Dưới đây là ý tưởng phương pháp đánh chỉ số động dựa trên mức độ đánh trọng của người sử dụng như sau:



Hình 12 : Một ví dụ trong phương pháp đánh chỉ số

Từ hình trên ta miêu tả phương pháp đánh trọng số như sau, khi người dùng nhập vào một tính huống vào cơ sở tình huống, thuộc tính có trọng số cao nhất sẽ được chọn làm chỉ số cho việc rút trích. Thuộc tính đó sẽ làm mờ trở thành biến ngôn ngữ, giả sử trong hình trên là 3 giá trị “Cao”, “TB”, “Thấp”. Thì nếu giá trị mờ của thuộc tính đó được mờ hóa thuộc biến nào ở một trong 3 biến đó thì sẽ được đưa vào khu vực chứa các thuộc tính này. Ví dụ, nếu biến mờ là “Cao”. Trong ví dụ, là EPS giả sử là cao. Thì nó sẽ được đưa vào khu vực $\text{Index}(\text{EPS} + \text{“Cao”})$. Ý nghĩa công thức này là đánh chỉ số trên thuộc tính là EPS và giá trị là “Cao”.

3.1.1.1.5. Rút trích tình huống

Phương pháp k láng giềng gần nhất (k-NN)

Như trong biểu đồ hình 11, có 2 phương pháp được sử dụng trong quá trình rút trích tình huống: Mạng nơron và k-NN trong đó k-NN sử dụng 2 độ đo khác nhau là Manhattan và Euclidean.

Phương pháp k-NN được đề xuất như sau:

B1. Thiết lập các độ đo khoảng cách như: Manhattan, Euclidean.

B2. Dựa vào phương pháp đánh chỉ số động ở trên ta thực hiện rút trích k-neighbor gần nhất dựa trên các độ đo khác nhau mà ta đã chọn lựa ở trên.

B3. Với k-neighbour gần nhất của mỗi phương pháp tính khoảng cách xem xét kết quả của nó.

Xét tỉ lệ: $\delta = \max(\alpha, \beta, \gamma)$

α : Số tình huống có kết quả "Tốt"

β : Số tình huống có kết quả "TB"

γ : Số tình huống có kết quả "Xấu"

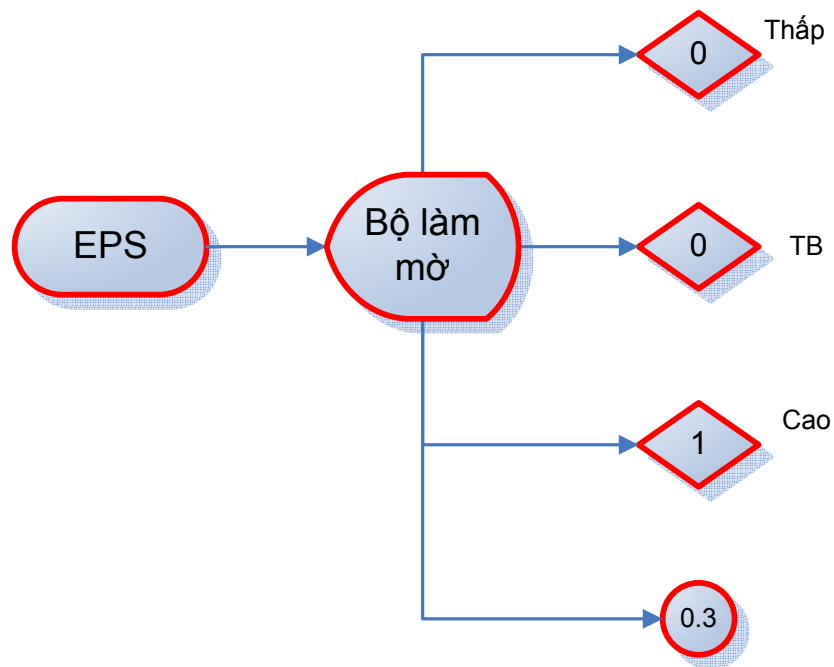
B4. Từ việc xác định giá trị δ lớn nhất chúng ta sẽ được kết quả dựa trên độ đo ta chọn.

Bảng 3 : Phương pháp rút trích k-NN

Phương pháp mạng nơron

Giả sử có một tình huống muốn đưa vào mạng nơron để huấn luyện hoặc phân loại tình huống. Trước hết sẽ được mờ hóa các thuộc tính biểu diễn như sau:

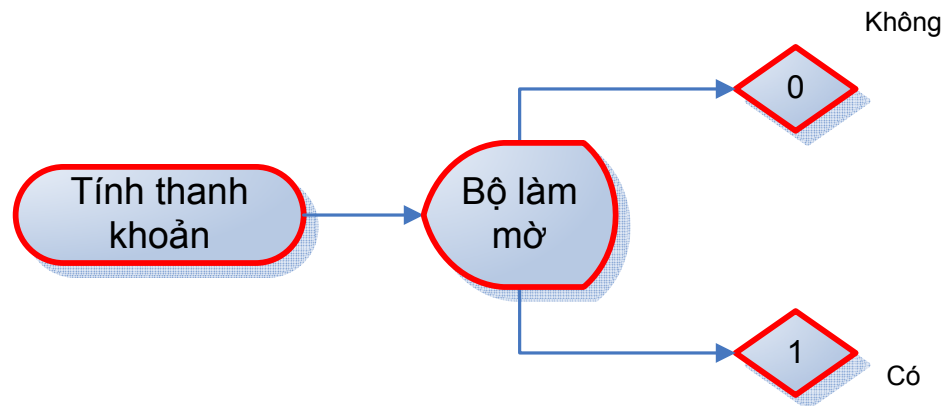
Giả sử với chỉ số EPS được đưa vào ta mờ hóa như hình dưới đây:



Hình 13 : Ví dụ về phương pháp mờ hóa thuộc tính EPS sử dụng cho đầu vào mạng nơron

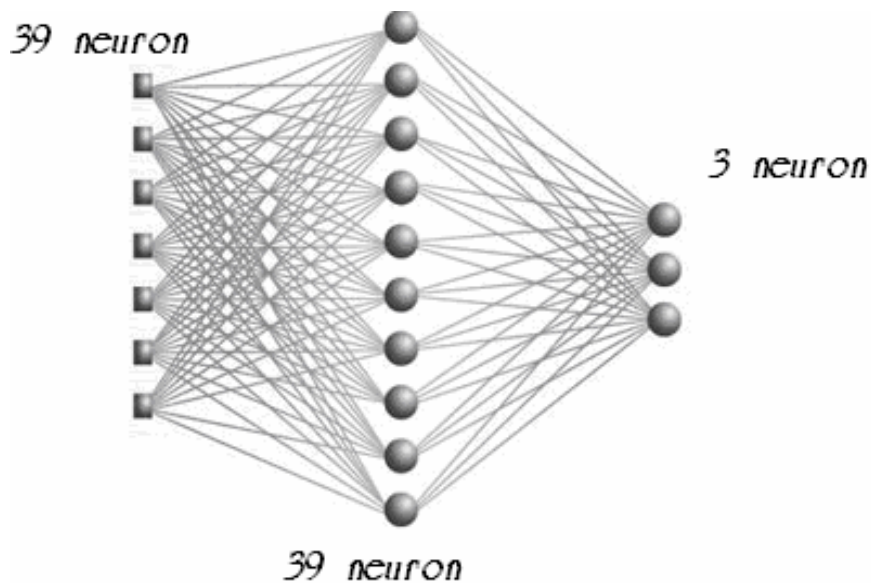
Giả sử EPS có giá trị là A. Khi đưa vào bộ làm mờ thuộc tính, nó sẽ được làm mờ với đầu ra gồm 4 phần. Giả sử trong hình 12 ở trên với giá trị A là có giá trị mờ là “Cao”. Thì phần đầu ra của ở “Cao” là 1 và hai phân khác là 0. Và giá trị phụ thuộc vào giá trị ”Cao” là 0.3.

Làm tương tự với tất cả các thuộc tính khác tùy vào những qui định đã trình bày ở bảng giá trị ở trên mà ta xây dựng ra một bộ các giá trị đầu ra cho mỗi biến. Tuy nhiên một số các thuộc tính trong bảng 1 vì giá trị đầu vào ngay ban đầu là các biến ngôn ngữ nên ở đây có hơi khác trong biến đổi so với các thuộc tính trước đó. Ví dụ, sơ đồ thể hiện như sau:



Hình 14 : Bộ làm mờ biến ngôn ngữ

Như vậy, sau quá trình làm mờ thì ta sử dụng tổng tất cả các giá trị đầu ra của tất cả các thuộc tính. Như vậy với số thuộc tính mà ta sử dụng là 12 mỗi thuộc tính có thể có 2 hoặc 3 giá trị mờ. Theo tính toán thì số biến đầu ra sau quá trình làm mờ là 39. Vậy ta sẽ xây dựng mạng nơron gồm 39 giá trị đầu vào như sau:

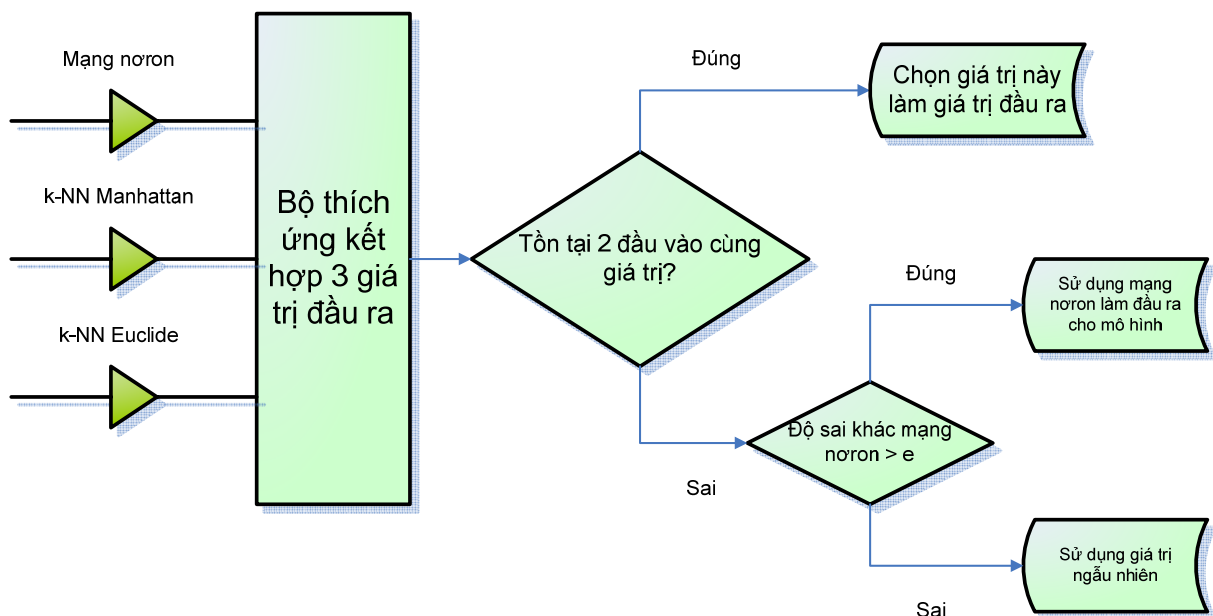


Hình 15 : Cấu trúc mạng nơron nhằm phân loại tình huống sau khi làm mờ

Số nơon lớp ẩn được sử dụng bằng số nơon lớp nhập và giá trị đầu ra là 3 nơon thể hiện cho 3 lớp đối tượng khác nhau cho kết quả của mỗi tình huống. Việc lựa chọn kết quả cuối cùng phụ thuộc vào giá trị đầu ra. Số nơon lớp ẩn chỉ là một đề xuất chứ ở đây chưa thực sự thực hiện quá trình kiểm chứng nào để chứng minh rằng việc chọn lựa này là tốt.

3.1.1.1.6. Thích nghi tình huống

Trong sơ đồ hình 11, mô hình thích nghi sẽ dựa vào ba mô hình mạng nơon và mô hình k-NN với 2 độ đo là Mahattan và Euclide. Mô hình thích ứng chi tiết như sau:



Hình 16 : Mô hình thích nghi tình huống

Giải thích sơ đồ:

Sau quá trình rút trích cho phân loại trong mạng nơon và k-NN thì những kết quả của những giá trị đầu ra sẽ được đưa vào bộ thích ứng, bộ thích ứng sẽ dựa vào các giá trị này để đưa ra quyết định giá trị đầu ra ứng dụng của hệ thống.

Có 2 trường hợp xảy ra với phương pháp này:

- Trường hợp nếu có 2 đầu ra cùng một giá trị thì giá trị đầu ra cuối cùng sẽ chọn giá trị này
- Ngược lại sẽ xảy ra hai trường hợp:
 - Nếu độ sai khác trong mô hình mạng nơron giữa các giá trị đầu ra là lớn $> \xi$. Thì sẽ sử dụng giá trị của mạng làm giá trị đầu ra.
 - Ngược lại, thì có thể chọn ngẫu nhiên giá trị của 2 kết quả còn lại làm giá trị cho đầu ra.(Hoặc nếu chần chẫn hơn thì có thể gửi thông điệp là bài toán không có lời giải).

3.1.1.1.7. Bảo trì cơ sở tình huống

Việc bảo trì cơ sở tình huống luôn là một vấn đề quan trọng, nhằm đảm bảo an toàn, ổn định và chính xác. Có khá nhiều phương pháp nhưng ở đây sử dụng một số các phương pháp sau cho việc bảo trì:

Loại dữ liệu thừa: Trong mô hình dự đánh giá chứng khoán này mục đích phân loại một tình huống thuộc vào một trong 3 tình lớp “Tốt”, “TB”, “Xấu”. Phương pháp áp dụng ở đây đơn giản như sau: Tìm tất cả các giống nhau ở tất cả các thuộc tính. Giữ lại một trong các tình huống đó, còn lại là loại bỏ tất cả các tình huống tương tự với nó.

Loại bỏ dữ liệu đựng độ: Trong một số trường hợp nếu tồn tại 2 tình huống mà giống nhau ở tất cả các thuộc tính nhưng lại cho những kết quả khác nhau. Điều này cũng cần được xem xét, trong việc loại tình huống nào và chọn tình huống nào? Trong khóa luận chỉ sử dụng một phương pháp đơn giản là chọn ngẫu nhiên một trong những tình huống đó và loại đi. Nếu có nhiều trường hợp đựng độ hơn thì việc chọn những giá trị nào có số tình huống đựng độ ở kết quả cao nhất. Sẽ chọn giữ lại một trong chúng và những tình huống đựng độ khác sẽ loại bỏ đi. Ví dụ, Giả sử có 3 tình huống đựng độ,

trong đó có 2 tình huống có cùng giá trị kết quả A còn lại tình huống kia là B. Thì ta sử dụng A làm kết quả và loại đi 2 tình huống kia.

3.2 Áp dụng trong phân tích kỹ thuật

3.2.1. Mô hình mạng nơron

Dự đoán là một quá trình tạo các giá trị đầu ra qua tập các biến đầu vào, các biến thường là dữ liệu trong quá khứ. Nói một cách cơ bản, dự đoán giả sử các giá trị xảy ra dựa trên ít nhất một phần các sự kiện quan sát đã xảy ra trong quá khứ. Những mối quan hệ trong quá khứ có thể được khám phá qua quá trình nghiên cứu và quan sát. Ý tưởng cơ bản của dự đoán là quá trình tìm một xấp xỉ ánh xạ giữa giá trị đầu vào và giá trị đầu ra theo một tuần tự để khám phá các luật không rõ ràng liên quan đến những chuyển động được quan sát. Ví dụ, dự đoán giá chứng khoán có thể được miêu tả theo những cách sau. Giả sử u_i là giá chứng khoán hiện tại, v_i là giá chứng khoán 10 ngày tiếp theo. Nếu sự dự đoán giá chứng khoán sau 10 ngày có thể đạt được sử dụng giá trị chứng khoán của hôm nay thì có một hàm ánh xạ giữa u_i và v_i , trong đó $u_i = F_i(v_i)$ sử dụng tất cả các cặp (u_i, v_i) dữ liệu trong quá khứ, một hàm F chung bao gồm các hàm F_i có thể đạt được $v = F(u)$. Nói chung, \bar{u} bao gồm nhiều thông tin của giá hôm nay có thể sử dụng trong hàm F . Trong mạng nơron việc xác định hàm F sao cho chính xác phụ thuộc vào cách thức huấn luyện của mạng.

Trong các ứng dụng dự đoán tài chính đòi hỏi lượng dữ liệu khá nhiều thường thì khoảng 7-10 năm có thể nhiều hơn. Để xây dựng được một mô hình dự đoán tài chính tốt, bền vững và ổn định chúng ta cần phải chú ý đến một số vấn đề sau:

Thứ nhất, thay vì chỉ tập trung nhấn mạnh vào việc dự đoán giá chính xác, các tiêu chuẩn khác cũng cần được xem xét, hiện các nhà nghiên cứu về vấn đề dự đoán tài chính thường tập trung vào sử dụng tiêu độ thích nghi hay tiêu chuẩn tương tự nhằm xây dựng các mô hình trong lĩnh vực tài chính. Theo lý thuyết thì những nghiên cứu này là một phương pháp đúng. Như chúng ta hiểu thì trong thực tế không có một mô

hình dự đoán nào là hoàn hảo, chẳng có một mô hình dự đoán nào là đạt được mục đích một cách lý tưởng, do vậy việc tìm kiếm một kỹ thuật dự đoán hoàn hảo hiện nay là nằm ngoài khả năng, chúng ta chỉ có thể tối ưu mô hình dự đoán không hoàn hảo và sử dụng các tiêu chuẩn so sánh để đưa ra được một độ đo mang tính thực tế nhất.

Thứ hai, chúng ta nên có một phương pháp xử lý dữ liệu dự đoán một cách hợp lý. Việc tiền xử lý và lấy mẫu cho dữ liệu đầu vào có thể ảnh hưởng lớn đến kết quả thực hiện dự đoán của chúng ta. Việc chọn lựa dữ liệu đầu vào qua việc phân tích tính toán và một phần mang tính cảm giác có thể giúp chúng ta loại bỏ được những dữ liệu dư thừa. Hơn nữa kết quả dự đoán của mạng nơron có thể hiệu quả hơn nhiều. Ví dụ, khi kết quả so sánh là không hoàn hảo, chúng ta nên so sánh với mạng nơron đầu ra với dự đoán trước đó hoặc với dữ liệu thực khi các cấp độ giá được sử dụng như là các đích cho việc dự đoán.

Thứ ba, trong các hệ thống giao dịch thì mạng nơron không phải là một công cụ duy nhất cho dự đoán tài chính. Chúng ta cũng không nhắm đến công cụ nào là tốt nhất cho dự đoán. Trong thực tế, người ta không quan tâm nhiều về việc loại chuỗi thời gian (time-series) nào là thích hợp nhất cho mạng nơron mà phân tích dự đoán cuối cùng cho phép chúng ta hiểu được sự thích hợp của mô hình và của các chuỗi (series). Cuối cùng chúng ta kết luận rằng loại mô hình chắc chắn nào được sử dụng với loại chuỗi thời gian nào.

Thay vì chỉ thể hiện một thí nghiệm thành công hay mức độ tin cậy cho giá trị đầu ra của mạng. Dữ liệu nên được phân chia thành các phần nhỏ để tìm dữ liệu những tri thức đặc biệt trong chuỗi thời gian này. Thật quan trọng và cần thiết để nhấn mạnh 3 vấn đề này ở đây. Tiêu chuẩn khác nhau đối với học thuật (academic) và trong công nghiệp (industry). Trong học thuật, thỉnh thoảng người ta tìm kiếm độ chính xác 100% trong khi trong thực tế công nghiệp thực tế người ta hướng đến một kết quả chính xác nhưng bảo đảm chỉ ở 60%. Để xây dựng một mạng nơron hiệu quả chúng ta cần những thí nghiệm hiệu quả. Việc kiểm tra chỉ một thị trường và trong một khoảng thời gian

nhất định thì chẳng có ý nghĩa gì nhiều nó sẽ không dẫn đến một mô hình bền vững, Một mô hình bền vững hơn cần không chỉ trên một thị trường và trong nhiều khoảng thời gian.

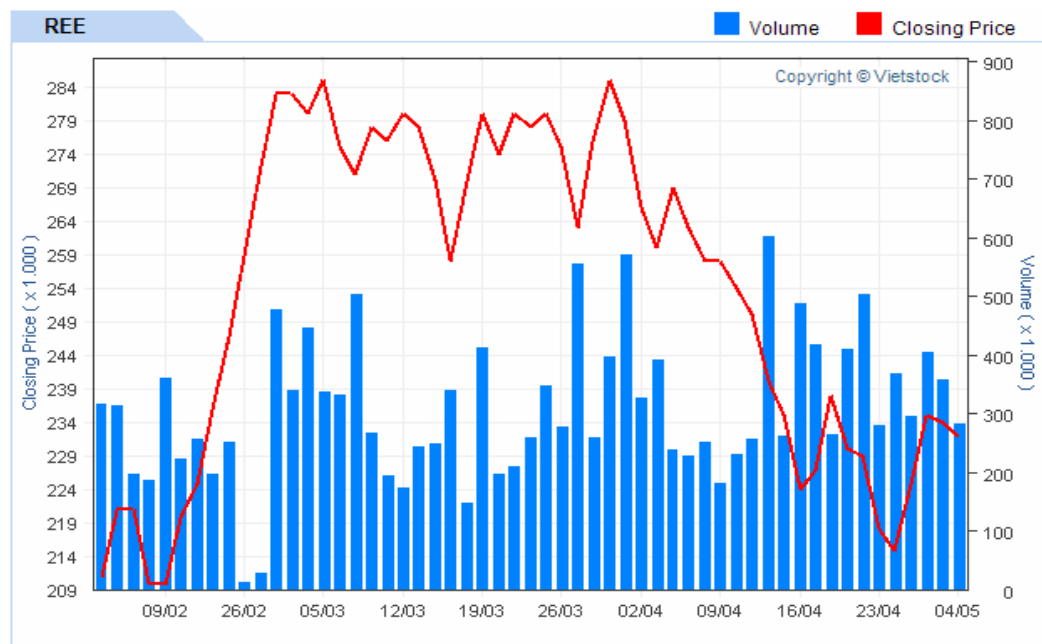
Một điều lưu ý là trong phương pháp này, tất cả các mạng sẽ được dùng là mạng nơron truyền thẳng 3 lớp với luật học delta.

3.2.1.1. Dự đoán dùng mô hình ARIMA

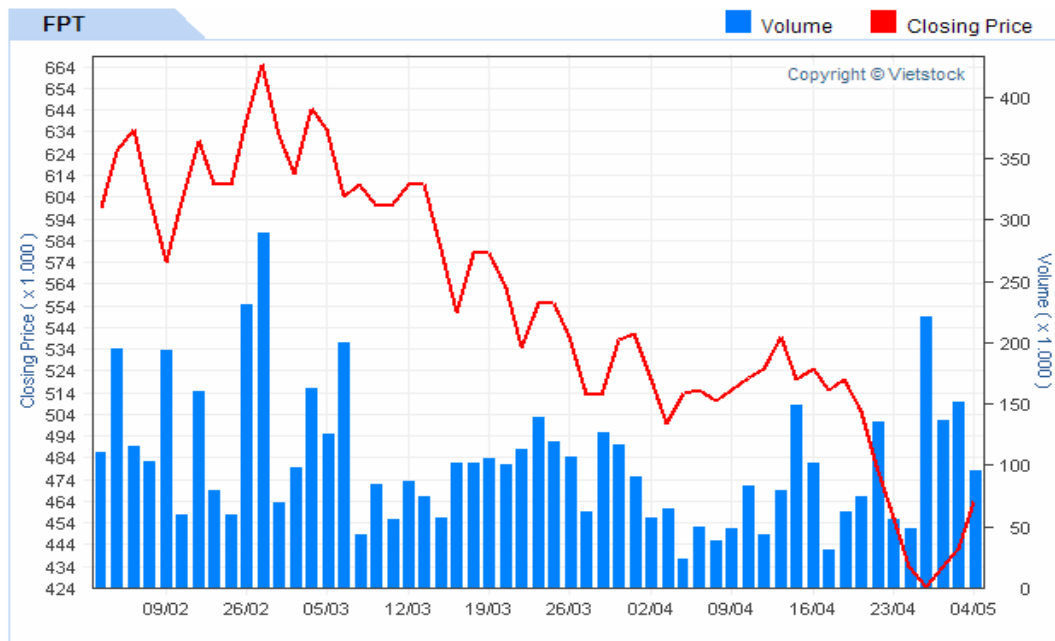
Một trong những kĩ thuật mà thường được sử dụng khá nhiều là phương pháp ARIMA, có thể xem thêm phương pháp ở phần phụ lục. Dưới đây ta sẽ đi vào chi tiết các kĩ thuật được áp dụng cho mô hình.

Lựa chọn dữ liệu (Data Selection)

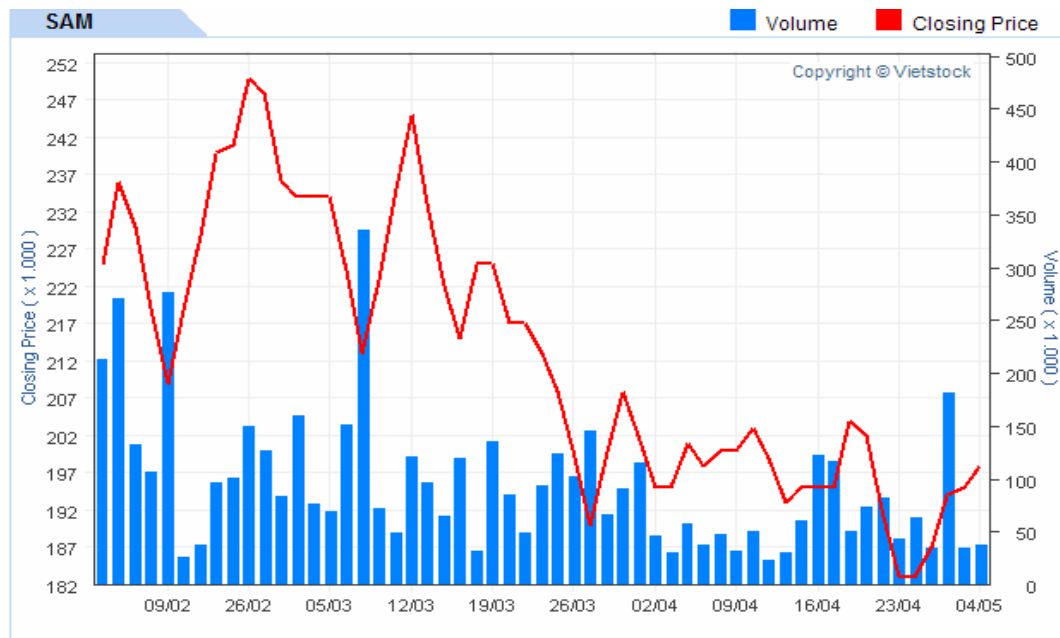
Vì thị trường chứng khoán Việt Nam so với nhiều nước trên thế giới là còn rất trẻ chỉ mới 7 năm trở lại đây và có những biến động khá thất thường ngoài ra còn phải điều chỉnh và sự ảnh hưởng nhiều của khá nhiều nhân tố khác. Do vậy việc chọn lựa được một bộ dữ liệu phù hợp và thật hợp lý không phải là điều đơn giản. Trong dữ liệu được sử dụng cho quá trình dự đoán trong mạng nơron được áp dụng ở đây chỉ đơn thuần là dự đoán giá cả của những ngày sắp tới. Vì vậy dữ liệu là toàn bộ giá đóng cửa của tất cả các cổ phiếu trên thị trường từ trước đến nay. Trong phương thức dự đoán giá cổ phiếu thường được sử dụng nhiều nhất chính là giá đóng cửa trong ngày của một cổ phiếu, tuy nhiên vì một số giới hạn nên ở đây chỉ phân tích và dự đoán giá chứng khoán trên sàn giao dịch chứng khoán thành phố.



Hình 17 : Giá giao dịch cơ phiếu REE (05/02/2007 – 05/05/2007)



Hình 18 : Giá giao dịch cơ phiếu FPT (05/02/2007 – 05/05/2007)



Hình 19 : Giá giao dịch cơ phiếu SAM(05/02/2007 – 05/05/2007)

Ngay từ định hướng cho vấn đề dự đoán là dựa đoán trên tất cả các loại cổ phiếu trên sàn giao dịch thành phố nên vấn đề dữ liệu không phải là vấn đề đơn giản. Trong ứng dụng xây dựng tất cả các dữ liệu được lấy từ trang web www.bsc.com.vn, www.hsc.com.vn ngoài ra dữ liệu cũng được lấy từ một số các trang khác như www.vnexpress.net hoặc www.vietstock.com.

3.2.1.1.1. Tiền xử lý dữ liệu (Data preprocessing)

Trước khi dữ liệu được phân tích thì vấn đề tiền xử lý cơ bản của dữ liệu là cần thiết, trong những ngày không giao dịch thì dữ liệu nhờ cũng sẽ được lấp đầy nhờ bước này nhờ một số kĩ thuật, cũng như quá trình này cũng nhằm giảm được tỉ lệ nhiễu, nhờ vậy sẽ giúp khả năng dự đoán của chúng ta được chính xác hơn. Đối với những nước có thị trường chứng khoán phát triển lâu và ổn định thì dữ liệu sẽ có tỉ lệ nhiễu thấp hơn so với dữ liệu của những nước mà thị trường chứng khoán đang phát triển như ở Việt Nam. Vì vậy việc tiền xử lý dữ liệu càng đóng một vai trò cực kì đáng kể

trong cả mô hình dự đoán. Một trong những kỹ thuật khử nhiễu được dùng nhiều nhất là phương pháp trung bình chuyển động (moving average). Phương pháp đơn giản như sau:

Cho chuỗi dữ liệu giá chứng khoán (time series) $\dots x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}, x_t \dots$ trở thành chuỗi dữ liệu sau :

$$\dots [x_{t-4} + x_{t-3} + x_{t-2}] / 3, [x_{t-3} + x_{t-2} + x_{t-1}] / 3, [x_{t-2} + x_{t-1} + x_t] / 3 \dots$$

Ngoài phương pháp khử nhiễu thì vấn đề không ổn định dữ liệu cũng là một vấn đề cần được xử lý trong bước này, theo thời gian thì dữ liệu có thể tăng hoặc giảm với mức độ sai khác rất lớn. Do đó nếu lấy những dữ liệu chuẩn theo giá của cổ phiếu (mặc dù đã khử nhiễu) thì kết quả nhận dạng sẽ không được chính xác. Để giải quyết vấn đề này người ta thường dựa vào độ biến động giá của cổ phiếu qua các ngày chứ không dựa vào giá cổ phiếu thực trên thị trường lúc đó.

Như trình bày ở trên, dữ liệu được sử dụng trong mạng nơron này là giá đóng cửa hằng ngày của cổ phiếu cần dự đoán. Tức là dựa trên chuỗi giá đóng cửa trong quá khứ ta sẽ dự đoán giá đóng cửa trong ngày hôm sau.

3.2.1.1.2. Biến đổi dữ liệu (Data Scaling)

Ở bước này ta thực hiện chuẩn hóa các giá trị đầu vào thành các giá trị $[0;1]$ hoặc $[-1;1]$ các giá trị này mới phù hợp cho giá trị đầu vào thường dùng trong mô hình mạng nơron. Việc lựa chọn phương pháp chuẩn hóa hợp lý cũng ảnh hưởng cực kì lớn đến kết quả dự đoán của chúng ta. Do đó, chọn lựa phương pháp chuẩn hóa như thế nào cũng cần được xem xét một cách kĩ càng. Trong quá trình thực hiện tìm hiểu, đã áp dụng một số những kỹ thuật ARIMA đơn giản. Nhưng thực sự khi áp dụng vào bài toán thì lại không đạt được một hiệu quả nhất định. Ngoài phương pháp ARIMA, còn tìm hiểu rất nhiều phương pháp khác nhau nhưng kết quả cuối cùng vẫn chưa thể so được với các bài báo nghiên cứu cùng một kỹ thuật này. Nhưng cuối cùng trong quá trình tìm hiểu đã đề xuất công thức sau đạt được hiệu quả cao nhất:

$$\mu = \frac{x_N - x_1}{N - 1}$$

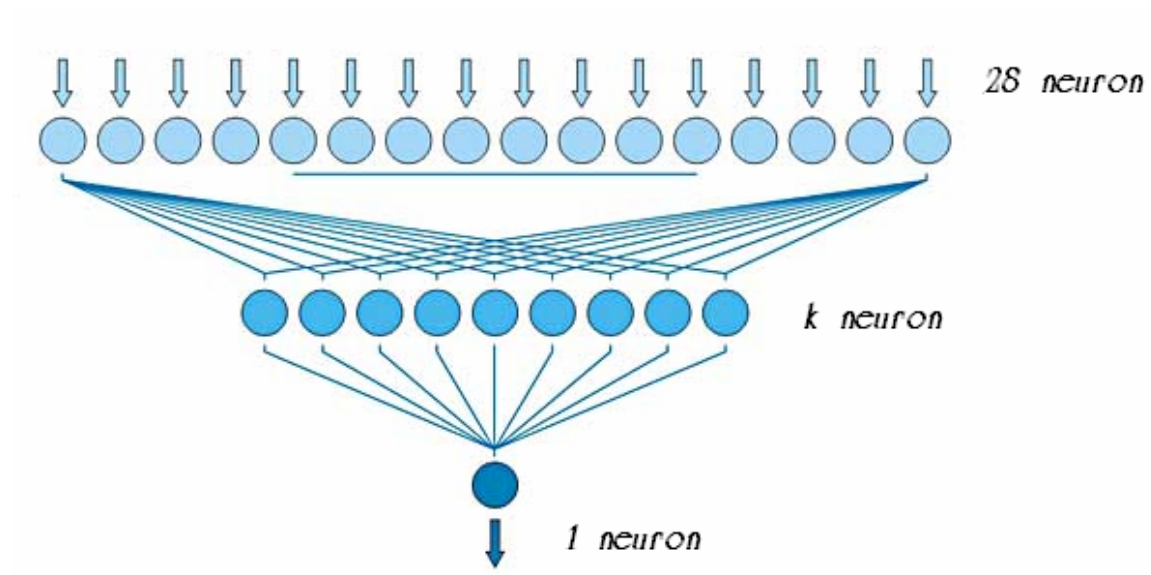
$$\sigma^2 = \frac{1}{N - 1} \left(\sum_{i=1}^{N-1} (x_{i+1} - x_i - \mu)^2 \right)$$

$$z_i = \frac{1}{1 + e^{-\left(\frac{x_{i+1} - x_i - \mu}{\sigma}\right)}}$$

Đây là công thức dựa trên nền tảng của mô hình ARIMA phát triển và đề xuất công thức, khi áp dụng vào bài toán thì kết quả là tốt hơn so với những mô hình khác. Có thể có những phương pháp khác nhằm chuẩn hóa dữ liệu hiệu quả hơn và chưa được tìm hiểu cũng như công bố. Vì vậy trong những công thức được sử dụng chỉ cần chọn một phương pháp mà ở đó áp dụng vào bài toán là tốt nhất. Và công thức trên là một sự lựa chọn.

3.2.1.1.3. Cấu trúc mạng nơron

Mạng được sử dụng là mạng 3 lớp lan truyền ngược (feedforward backpropagation), tuy nhiên khác với mạng thông thường ta thường sử dụng trong các bài toán nhận dạng, mạng sử dụng trong dự đoán chứng khoán có một chút khác biệt. Ở lớp xuất chỉ có một nơron duy nhất biểu diễn giá trị đầu ra của mạng đây cũng chính là giá đóng cửa dự đoán xem như được chuẩn hóa của ngày hôm sau. Chi tiết cấu trúc mạng qua hình sau:



Hình 20 : Mô hình mạng nơron sử dụng ARIMA

Cấu trúc mạng được sử dụng gồm 28 giá nơron tầng nhập. Khi thực hiện kiểm nghiệm như sau:

- Cho n là số nơron đầu vào. Sau đó cho chạy giá trị n từ 7 đến 50, trong đó số nơron tầng ẩn bằng số nơron tầng nhập. Thì kết quả đạt được ở $n = 28$ là đạt được hiệu quả tốt nhất. Tiếp tục thử kiểm tra trên một số giá trị n lớn hơn thì kết quả nhận được chỉ ra rằng với n càng lớn thì mức độ chính xác trong dự đoán càng giảm.
- Sau đó, cố định $n = 28$ và thay đổi số nơron tầng ẩn thì kết quả cuối cùng $k = 60$ là một lựa chọn đạt được hệ số lỗi thấp nhất. Đó là lý do tại sao lại chọn cấu trúc mạng trên.

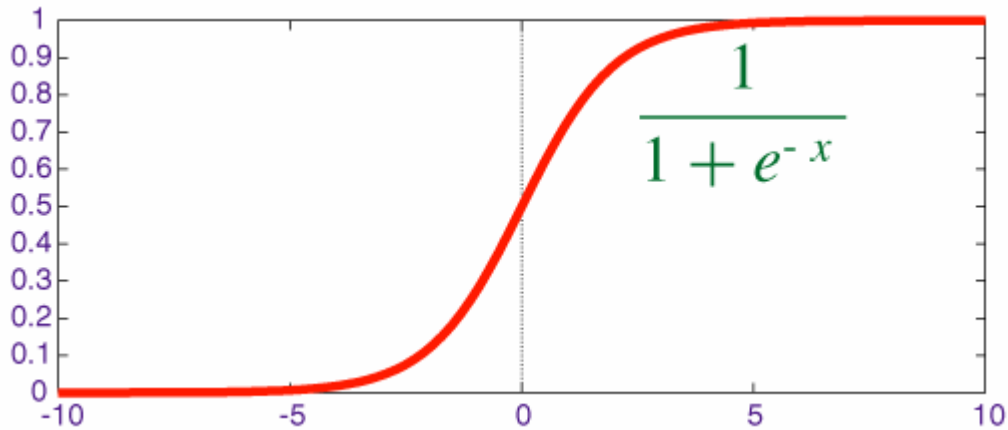
Vấn đề về mạng nơron chi tiết sẽ không bàn mà chỉ tập trung trên một số những vấn đề cốt yếu nhằm xây dựng mạng.

$$O_c = h_{Hidden} \left(\sum_{p=1}^P i_{c,p} w_{c,p} + b_c \right)$$

$$h_{Hidden}(x) = \frac{1}{1 + e^{-x}}$$

c,p: Số neuron ở lớp hiện hành (current), lớp trước đó (previous)

Khi mạng thực thi thì lớp ẩn sẽ thực hiện tính toán như biểu thức trên và đưa kết quả O_c đến lớp xuất. Trong đó $i_{c,p}$ là giá trị đầu vào từ lớp trước đó, đến lớp hiện hành và $w_{c,p}$ là trọng số kết nối giữa hai lớp trên, cuối cùng b_c là bias. Hàm h_{Hidden} là hàm sigmoid, có rất nhiều hàm khác cũng phù hợp cho bài toán nhưng ở đây sử dụng hàm này trong mô hình mạng cho bài toán dự đoán. Đồ thị hàm sigmoid đơn giản như dưới đây:



Hình 21 : Hàm sigmoid

Ở lớp xuất giá trị đầu ra cuối cùng O_c được tính toán bằng công thức:

$$O_c = h_{Output} \left(\sum_{p=1}^P i_{c,p} w_{c,p} + b_c \right)$$

$$h_{Output}(x) = x$$

3.2.1.1.4. Huấn luyện mạng

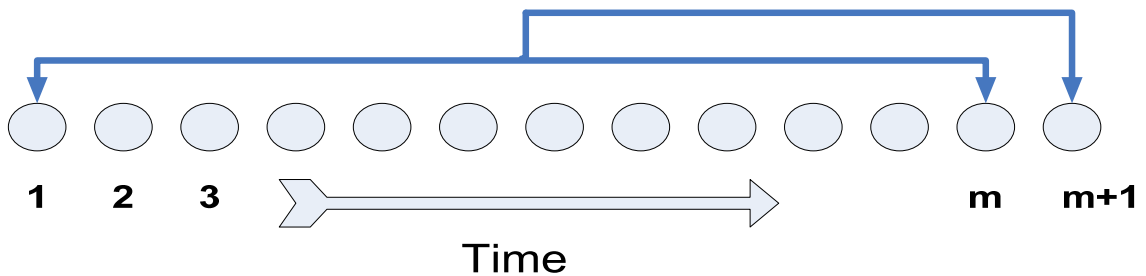
Huấn luyện mạng được sử dụng dựa trên luật học Delta mà ta đã biết nên ở đây không bàn lại vấn đề này, mà sẽ đi chi tiết phương pháp chọn bộ mẫu huấn luyện mạng.

Cho chuỗi n chuỗi giá trị chứng khoán (giá đóng cửa) theo thời gian (bộ mẫu học) như sau $y_{i_1}, y_{i_2}, \dots, y_{i_{m+1}}$. Mỗi chuỗi là một quan sát trong một khoảng thời gian của một loại chứng khoán. Mạng sẽ thực hiện kỹ thuật học với m giá trị trước đó sẽ dự đoán được giá trị của ngày tiếp theo:

$$\begin{aligned} y_{1_{m+1}} &= f(y_{1_1}, y_{1_2}, \dots, y_{1_m}) \\ y_{2_{m+1}} &= f(y_{2_1}, y_{2_2}, \dots, y_{2_m}) \\ &\dots \\ y_{n_{m+1}} &= f(y_{n_1}, y_{n_2}, \dots, y_{n_m}) \end{aligned}$$

Phương pháp học để tìm trọng số kết nối và mức độ lỗi của mạng trong quá trình học.

Sau khi học sẽ thực hiện quá trình test các bộ mẫu test mà chúng ta chọn sẵn với cùng phương pháp trên. Với m giá trị đầu vào đưa ra giá trị dự đoán tiếp theo, giá trị thứ m+1. Tức là quá trình huấn luyện và dự đoán sẽ dựa vào những kết quả của m = 28 để dự đoán giá đóng cửa của ngày hôm sau.



Hình 22 : Phương pháp áp dụng chuỗi giá vào mạng nơron

Bộ dữ liệu học sẽ được sử dụng là 2/3 toàn bộ dữ liệu thị trường của một cổ phiếu từ khi bắt đầu, và 1/3 dữ liệu sẽ được sử dụng để kiểm tra. Với các phương pháp tính lỗi được trình bày ở phần dưới đây.

3.2.1.1.5. Phương thức đánh giá dự đoán

Với mục đích đánh giá khả năng dự đoán của mạng trên bộ mẫu kiểm tra, chúng ta khảo sát mức độ chính xác bằng cách tính 3 kiểu đánh giá thống kê sau: the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), đây là 3 phương pháp được sử dụng nhiều trong hầu hết các hệ thống dự đoán giá chứng khoán nói riêng và dự đoán tài chính nói chung. Công thức như sau:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - A_i)^2}{n}}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - A_i|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - A_i}{A_i} \right| * 100$$

A_i : Giá trị thực.

P_i : Giá trị dự đoán.

n : Chiều dài của chuỗi giá.

Trong 3 hệ số này thì hệ số MAPE thường được đánh giá cao bởi tính ổn định không phụ thuộc vào khoảng thời gian kiểm tra. Ở khóa luận cũng sẽ đánh giá chính dựa trên tiêu chí này.

3.2.1.2. Dự đoán dùng các chỉ số kỹ thuật

Trước khi kết thúc phần trình bày về các phương pháp và thuật toán áp dụng mạng nơron nhằm giải quyết bài toán. Một phương pháp tiếp cận hơi khác so với thuật toán trước, tuy vẫn sử dụng mạng nơron nhưng có một số khác biệt sau:

- P** *Phương pháp chỉ sử dụng các chỉ số kỹ thuật như: RSI, OBV v.v... Có thể xem kỹ những chỉ số này ở phần phụ lục.*
- P** *Không dự đoán ngày tiếp theo mà thay vào đó dự đoán ngày thứ 4 trong tương lai.*

Đặc trưng đầu vào của mạng

Các chỉ số này được trình bày một cách chi tiết về chức năng và cách sử dụng các chỉ số này trong phân tích kỹ thuật ở phần phụ lục. Đây là một trong những chỉ số được sử dụng phổ biến nhất.

Open

Close

Volume

Accumulation/Distribution

RSI

Stock%D

Stock%K

%R

Arms Ease of Movement

OBV

VolPriceTrend

ChangeIn4Close

Bảng 4 : Dữ liệu ứng dụng cho mạng nơron trong phân tích kỹ thuật

Open: Giá mở cửa trong ngày

Close: Giá đóng cửa cuối ngày

Volume: Số lượng giao dịch của cổ phiếu trong ngày.

Accumulation/Distribution, RSI, Stock%D, Stock%K, %R, AEM, OBV, VPT: Là các chỉ số thường được sử dụng trong phân tích kỹ thuật. Những chỉ số này sẽ được trình bày ở phần phụ lục.

ChangeIn4Close: là tỉ lệ phần trăm giá 4 ngày sau đó so với giá của ngày dự đoán hiện thời. Đây chính là giá trị đầu ra duy nhất của ứng dụng.

Cấu trúc mạng

Mạng được sử dụng trong phương pháp này là mạng lan truyền ngược 3 lớp với các thông số với các hàm sử dụng như ở phương pháp trên.

Cấu trúc mạng bao gồm: (1) 11 nơon tần nhập, (2) 20 nơon tầng ẩn (3) 1 nơon tầng xuất. Phương pháp này có một chút khác so với mạng nơon sử dụng ở phương pháp trên là thay vì dự đoán giá đóng cửa của ngày hôm sau thì có thể dự đoán giá đóng cửa của n ngày tiếp theo. Và trong ứng dụng chọn $n = 4$.

Chuẩn hóa giá trị đầu vào

$$SF = \frac{(SRMax - SRMin)}{(XMax - XMin)}$$
$$\underline{Xp = SRMin + (X - XMin) * SF}$$

- Xp – giá trị thực của cột.
- $XMin$ – giá trị thực nhỏ nhất của cột.
- $XMax$ – giá trị thực lớn nhất của cột.
- $SRMin$ – giá trị giới hạn dưới của khoảng tỉ lệ.
- $SRMax$ – giá trị giới hạn trên của khoảng tỉ lệ.
- SF – nhân tố tỉ lệ.

- Xp – giá trị được chuẩn hóa.

3.2.2. Mô hình ECBR

Mạng nơron đã được sử dụng nhiều trong các ứng dụng dự đoán tài chính nói chung cho nhiều thập kỉ qua. Còn việc sử dụng CBR để dự đoán là lĩnh vực rất ít được sử dụng. Phần này sẽ đề xuất một phương pháp khác mạng nơron là kĩ thuật áp dụng CBR.

Vì sao lại sử dụng CBR ?

Hệ thống CBR là một hệ thống mà đưa ra quyết định bằng cách sử dụng các tri thức trong quá khứ, một hệ thống dự đoán nói chung chắc chắn chỉ có thể dựa trên dữ liệu quá khứ và một số các thuật toán được phát triển nhằm đưa ra những dự đoán những vấn đề ở tương lai, đây chính là điểm mạnh của một hệ CBR vì vốn bản thân nó sở hữu phương pháp này, nên lựa chọn nó cho quá trình dự đoán thì cũng là vấn đề hợp lý.

Lợi ích quan trọng nhất của CBR là tương tự khả năng học của con người. Con người quan sát, sử dụng những dữ liệu trong quá khứ nhằm đưa ra những quyết định của tương lai. Thường thì phép ngoại suy (extrapolation) cho một trường hợp mới là đặc biệt hoặc “lụi” (ad hoc). Trong các trường hợp khác thì phép ngoại suy mang tính hình thức hơn, và thường dựa vào các mệnh đề quy nạp như các công thức, các luật v.v...

3.2.2.1. Thể hiện case và đánh chỉ số case

Khác với các biểu diễn tình huống thông thường, mô hình tình huống trong mô hình dự đoán được xây dựng phần lớn là dựa vào giá chứng khoán qua một khoảng thời gian giả sử M ngày được biểu diễn như sau:

	Day 1	Day 2	Day 3	...	Day M
Open	O_1	O_2	O_3	...	O_M
High	H_1	H_2	H_3	...	H_M
Low	L_1	L_2	L_3	...	L_M
Close	C_1	C_2	C_3	...	C_M
Volume	V_1	V_2	V_3	...	V_M

Bảng 5 : Tình huống trong mô hình ECBR

Mỗi tình huống là M ngày sử dụng một số thông số thị trường như: **Open** (Giá mở cửa của cổ phiếu trong ngày), **High** (Giá cao nhất của cổ phiếu trong ngày), **Low** (Giá thấp nhất của cổ phiếu trong ngày), **Close** (Giá đóng cửa của cổ phiếu), **Volume** (Số lượng giao dịch trên cổ phiếu đó trong ngày).

3.2.2.2. Đánh chỉ số tình huống

Việc truy cập để rút trích nhanh và hiệu quả đối với một CBR thì việc đánh chỉ số như thế nào cho hợp lý sao cho khả năng lựa chọn tình huống rút trích đạt được kết quả càng cao càng tốt, và việc rút trích này cũng yêu cầu được sử dụng khách quan và phù hợp với thực tế. Một tình huống chính bản thân nó là động chứ không tĩnh như những phương pháp CBR truyền thống, việc mô tả một tình huống như trên chỉ mang tính thể hiện.

Khi xây dựng ứng dụng việc đánh chỉ số tình huống được sử dụng một trong 2 phương pháp sau:

Phương pháp 1: Thêm một số thuộc tính trong biểu diễn tình huống như trong bảng dưới đây, ta thêm 3 thuộc tính:

```

ECBRCase{
    MCK : Mã cổ phiếu.
    Date : Ngày thể hiện case.
    Vary7 : Độ biến động trong 7 ngày.
    Vary14 : Độ biến động trong 14 ngày.
    Vary21 : Độ biến động trong 21 ngày.
}

```

Bảng 6 : Bảng đánh chỉ số trong ECBR

Phương pháp biểu diễn tình huống này là hiệu quả và giải quyết được vấn đề ta đã nêu trên.

Phương pháp 2: Sử dụng viết rút trích ngẫu nhiên bộ học và bộ test. Tức là trong quá trình học hoặc test các bộ này được sử dụng ngẫu nhiên.

Việc biểu diễn linh động chiều dài tình huống cũng là một vấn đề cần bàn đến, không chỉ dựa vào một thời gian cụ thể mà sử dụng nhiều khoảng thời gian khác nhau điều này tăng tính tối ưu thêm cho hệ thống CBR khi thực hiện quyết định. Vì theo đánh giá bản thân thì nếu sử dụng cứng một khoảng thời gian nhất định giả sử như nếu ta chỉ sử dụng thời gian là 7 ngày để thực hiện dự đoán thì mỗi ngữ cảnh chứng khoán luôn thay đổi, việc sử dụng nhiều biến ngữ cảnh hơn tất nhiên hiệu quả đem lại sẽ cao hơn.

Ngày thể hiện tình huống: Là ngày cuối cùng của một tình huống truyền thống mà đã mô tả ở trên.

Độ biến động 7, 14, 21 ngày: chính là tính mức độ biến động của tình huống đó theo công thức sau:

$$\delta = \frac{1}{N} \sum_{i=1}^N |p(i) - p(i-1)|$$

3.2.2.3. Giải quyết vấn đề với CBR truyền thống

Trước khi đi vào trình bày thuật toán Ensemble CBR xin nhắc lại phương pháp CBR truyền thống sử dụng kết hợp k láng giềng gần nhất. Phương pháp truyền thống này thường tìm kiếm một thể hiện tốt nhất hoặc đánh trọng kết hợp giữa các một số nhỏ những người láng giềng trong không gian quan sát, nó tương tự phương pháp trong mạng nơron trong các bài toán nhận dạng mẫu. Thuật toán đơn giản như sau:

B1. Bắt đầu với case hiện hành $x(t)$

B2. Tìm k case gần nhất $x(t_i)$ trong quá khứ mà gần với $x(t)$ nhất theo hàm khoảng cách $d = d(x(t_i), x(t))$.

B3. Tính tổng trọng số

$$d_{TOT} = \sum_{i=0}^n d_i$$

B4. Xác định trọng số liên quan của láng giềng i th:

$$w_i = \frac{1}{j-1} \left[1 - \frac{d_i}{d_{TOT}} \right]$$

B5. Tìm kết quả $x(t_i + 1)$ của mỗi tình huống $x(t_i)$ trong tập các láng giềng.

B6. Tính dự đoán tại thời điểm $(t + 1)$:

$$\hat{x}(t+1) = \sum_{i=0}^j w_i x(t_i + 1)$$

Bảng 7 : Thuật toán CBR truyền thống trong dự đoán chứng khoán[39]

3.2.2.4. Giới thiệu phương pháp Ensemble CBR

Phương pháp kết hợp (Ensemble) gần đây đã trở thành một khóa luận nóng trong cộng đồng máy học và tính toán thông minh, phương pháp này đã được áp dụng khá thành công trong các hệ thống học như mạng nơron và cây quyết định. Đặc biệt ý

tưởng này đã được áp dụng thành công trong lĩnh vực nhận dạng chữ viết (Optical character recognition), Nhận dạng mặt người (face recognition), Phân tích ảnh khoa học (scientific image analysis) , Chuẩn đoán ảnh y khoa (medical diagnosis) v.v... Ý tưởng này bắt nguồn từ đầu thập kỉ 1990 nhằm dự đoán những vấn đề không ổn định, phương pháp này khi kết hợp với các hệ thống khác nhất là trong các hệ thống có khả năng học như mạng nơron ... đem lại những cải tiến mạnh mẽ và đạt được những hiệu quả cao hơn so với khi thực hiện hệ thống trên một cách độc lập. Trong thị trường chứng khoán thì vấn đề bất ổn, sự thay đổi liên tục là điều dễ nhận biết, việc sử dụng một hệ CBR đơn thuần không đem lại hiệu quả cao, xin đề xuất một phương pháp hoàn toàn mới trong việc sử dụng hệ CBR trên tư tưởng kết hợp với mong muốn đạt được hiệu quả tốt hơn.

3.2.2.5. Thuật toán Ensemble CBR

II Thuật toán gồm 3 phần chính:

- ☐ *Chọn lựa và chuẩn hóa dữ liệu*
- ☐ *Tối ưu mô hình.*
- ☐ *Dự đoán kết quả.*

Thuật toán được đề xuất ở đây là tương tự một phương pháp sử dụng “nhiều CBR” thông qua việc kết hợp các vector thuộc tính đưa ra một mô hình tối ưu nhằm dự đoán kết quả. Dựa vào sự kết hợp của các vector đặc trưng của k láng giềng gần nhất và chuỗi các vector ngữ cảnh của các tình huống hiện hành, sau đó là quá trình đánh giá dựa trên hệ số lỗi MAPE, RMSE, MAE v.v... Sau khi dự đoán kết quả tại thời điểm $(t + 1)$ bằng cách sử dụng cửa sổ đánh giá, nó đưa ra bộ số dự đoán và thích nghi tình huống hiện hành. Mô hình dự đoán được lựa chọn qua đánh giá dựa trên bộ mẫu học và bộ mẫu kiểm tra, vấn đề đánh giá là một vấn đề khác quan trọng trong các kĩ thuật học như trong mạng nơron và CBR. Ý tưởng quan trọng nhất của ECBR (Ensemble CBR) là dự đoán $t+1$ dựa trên việc kết hợp các vector đặc trưng và đánh

giá kết quả dựa vào các tiêu chuẩn như MAPE, RMSE v.v... nhằm chọn ra mô hình tối ưu trong giai đoạn kiểm tra. Sau đó nó áp dụng mô hình tối ưu được chọn ở bước trước nhằm dự đoán $t+1$ và thiết lập lại các cửa sổ đánh giá để nhận dạng các kết quả tiếp theo. Ở đây để tính toán các khoảng cách ta sử dụng 2 độ đo đơn giản Manhattan, Euclidean và chuỗi các vector ngữ cảnh từ tình huống hiện hành bằng cách thay đổi các cửa sổ kiểm tra. Thuật toán chi tiết thực hiện như sau:

B1. Thực hiện khảo sát và phân tích dữ liệu xác định lựa chọn các mẫu.

B2. Với mỗi case $X(t)$ {

$$E(t) = \ln X(t)$$

$$DE(t) = E(t) - E(t-1)$$

$$ZDE(t) = (DE(t) - m_x) / s_x$$

}

B3. Thực hiện chạy CBR cho dữ liệu test.

a. Chọn tiêu chuẩn đánh giá như MAPE, RMSE.

b. Chọn các cửa sổ học và cửa sổ test.

B4. Tìm mô hình CBR cho mỗi thời điểm test $l = h \dots t$

// t là kết thúc của cửa sổ test.

For $i = 1 \dots \#$ của phương_thức_khoảng_cách.

For $j = 1 \dots \#$ của $n_láng_giềng_gần_nhất$

For $k = 1 \dots \#$ của vector_ngữ_cảnh

For $l = h \dots t$ {

//Dự đoán $ZDE(l)$ sử dụng các biến i, j, k .

$ZDE(l) = CBR(i, j, k);$

//Tính toán kết quả đánh giá.

}

Chọn mô hình CBR tối ưu với các biến $i_{best}, j_{best}, k_{best}$.

B5. Thực hiện ECBR cho việc dự đoán $v = t + 1 \dots z$

```

For  $v = t + 1 \dots z$  {
     $ZDE(t + 1) = ECBR(i_{best}, j_{best}, k_{best})$ 
    //Chuyển đổi ngược kết quả
     $DE(t) = \sigma * ZDE(t) + \mu$ 
     $E(t) = DE(t) + E(t - 1)$ 
     $X(t) = \exp(E(t))$ 
    //Trở lại bước 3.b để thiết lập lại cửa sổ test.
    //Làm lại bước 4 để dự đoán giá trị (t+2)
}

```

B6. Phân tích kết quả ECBR.

Bảng 8 : Thuật toán ECBR

3.2.2.6. Áp dụng thuật toán đề xuất vào bài toán dự đoán chứng khoán

Dữ liệu: là chỉ các giá trị hằng ngày của tất cả các chứng khoán có mặt trên thị trường chứng khoán, trên sàn giao dịch ở Thành phố Hồ Chí Minh. Lẽ ra việc lựa chọn các “biến thô” trong việc áp dụng cho bài toán dự đoán là phức tạp hơn, dự tính ban đầu là lựa chọn 5 chỉ số sau đây để thực hiện dự đoán:

- **Giá mở cửa (Open Value):** Giá mở cửa trong ngày của một loại chứng khoán.
- **Giá cao nhất (Daily High):** Giá cao nhất trong ngày của chứng khoán.
- **Giá thấp nhất (Daily Low):** Giá thấp nhất trong ngày của chứng khoán.
- **Giá đóng cửa (Close Value):** Giá đóng cửa trong ngày của chứng khoán đó.
- **Khối lượng giao dịch (Daily Volume):** Tổng giao dịch hằng ngày của chứng khoán.

Tại sao lại chọn 5 chỉ số trên? Theo các nhà phân tích kỹ thuật thì đây là những số liệu quan trọng nhất trên thị trường, nó là nền tảng để xây dựng nên những chỉ số kỹ thuật khác. Trong hầu hết tất cả các chỉ số kỹ thuật được sử dụng hiện nay đều được xây dựng từ 5 số liệu này. Do vậy việc đưa vào ứng dụng bằng cách sử dụng 5 chỉ số này là chuyện bình thường và cần thiết.

Nhưng ở đây có một số khó khăn xảy ra: Đối với sàn chứng khoán thành phố thì một ngày có 3 lần khớp lệnh nên việc xác định chỉ số “Giá cao nhất” và “Giá thấp nhất” là không được khả quan lắm. Vì 1 trong 2 giá trị này có thể hoặc trùng với giá mở cửa và giá đóng cửa.

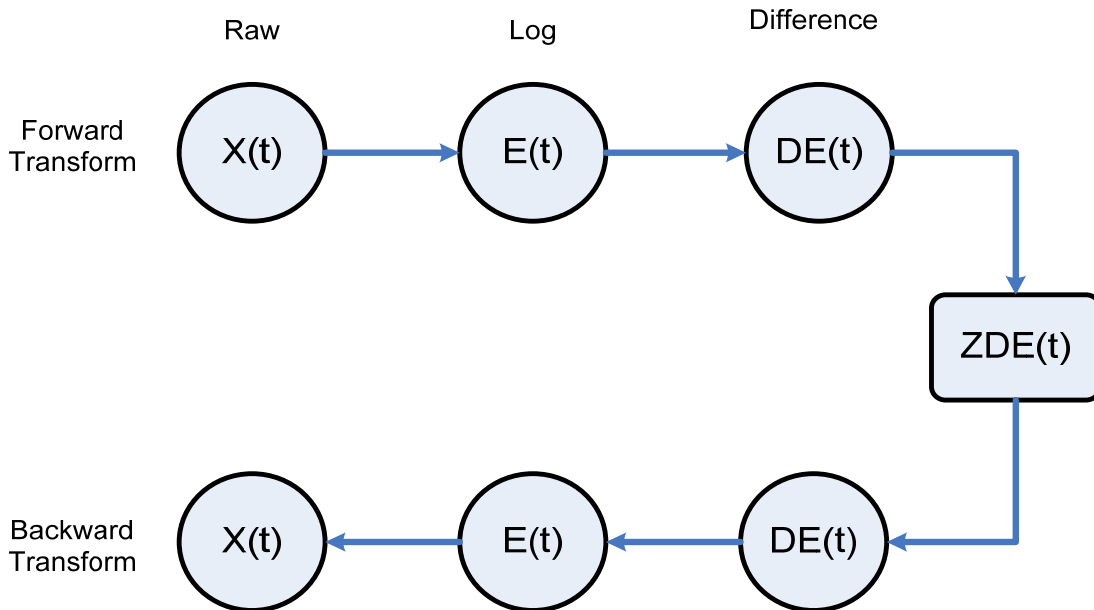
Vấn đề thứ 2 là trong những nghiên cứu liên quan ở một số nước trên thế giới thì những tác giả đó có rất nhiều thuận lợi:

- Thị trường phát triển một khoảng thời gian dài và ổn định.
- Có khối lượng dữ liệu lớn > 10 năm thuận lợi cho việc quan sát xác định những phân tích kỹ thuật cũng như nhìn nhận đánh giá một phần chu kỳ phát triển của thị trường.
- Đã có những người đi trước nghiên cứu trong lĩnh vực này và có một số kết quả nhất định, và họ dựa trên những nghiên cứu đó và đi tiếp.

Đối với thị trường Việt Nam nói chung và thị trường Thành Phố Hồ Chí Minh nói riêng những phát triển ban đầu còn nhiều xáo động. Ngay những ngày đầu luôn có những sự bất ổn và lượng cổ phiếu niêm yết trên thị trường lúc ban đầu là còn rất ít. Sự chuyển động lên xuống của giá cổ phiếu bị ảnh hưởng của nhiều nhân tố khách quan lẫn chủ quan, đối với những cổ phiếu mới niêm yết một vài năm gần đây thì lượng dữ liệu là chưa đủ để có thể thực hiện việc quan sát và dự đoán, và thực sự là rất hạn chế so với những nghiên cứu tương tự trên cùng lĩnh vực trên thế giới. Nói chung việc xây dựng ứng dụng dự đoán thị trường tốt không phải là một vấn đề đơn giản, tuy nhiên không phải là không thể làm được.

Chuẩn hóa dữ liệu

Một bước khá quan trọng khi xây dựng hệ thống nhận dạng, dự đoán. Giả sử trong trường hợp sử dụng những chỉ số như: Giá mở cửa, Giá cao nhất, v.v... Thì mỗi dữ liệu có một khoảng giá trị khác nhau và sự chênh lệch của nó cũng khác nhau. Nếu sử dụng những dữ liệu thô này đưa vào hệ thống thì thực sự là sẽ có rất nhiều vấn đề xảy ra và kết quả cuối cùng cũng theo chiều hướng xấu. Một phương pháp chuẩn hóa dữ liệu hiệu quả cũng có thể nâng cao xác suất dự đoán, cũng như khả năng nhận dạng của hệ thống.



Hình 23 : Sơ đồ chuẩn hóa ECBR

Tập hợp các công thức chuẩn hóa như sau:

Trước khi đi vào chi tiết công thức ta đặt 2 giá trị μ , σ sau bằng công thức :

$$\mu = \frac{1}{N-1} \sum_{i=1}^N \ln\left(\frac{x_i}{x_{i+1}}\right)$$
$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N \left(\ln\left(\frac{x_i}{x_{i+1}}\right) - \mu \right)^2$$

Áp dụng hai giá trị trên vào từng bước tương ứng trong hình 23 ta có 2 công thức chuyển đổi ngược như sau:

Forward Transform
$E(t) = \ln X(t)$
$DE(t) = E(t) - E(t-1)$
$ZDE(t) = \frac{(DE(t) - \mu)}{\sigma}$

Bảng 9 : Công thức chuyển đổi Forward Transform

Backward Transform
$DE(t) = \sigma * ZDE(t) + \mu$
$E(t) = DE(t) + E(t-1)$
$X(t) = \exp(E(t))$

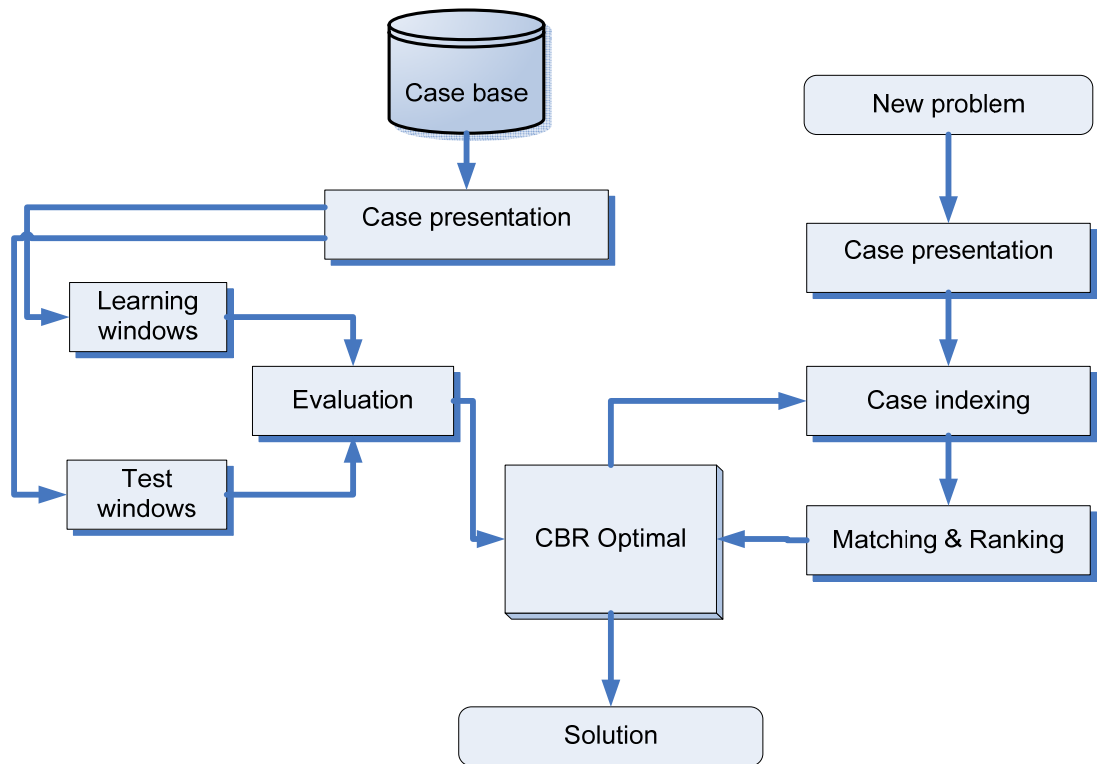
Bảng 10 : Công thức chuyển đổi Backward Transform

Trong đó:

- $X(t)$ là giá cổ phiếu ở thời điểm t cần được chuẩn hóa.
- $E(t)$, $DE(t)$, $ZDE(t)$ là từng bước chuẩn hóa sơ đồ chuẩn hóa hình 22.

Trên đây là một số trình bày sơ lược về thuật toán ECBR và là những ý tưởng sơ khởi về một loại phương pháp ứng dụng CBR vào dự đoán chứng khoán. Có thể kết quả đạt được khi xây dựng ứng dụng và mức độ hợp lý của nó như thế nào so với thực tế còn phải chờ một quá trình thử nghiệm lâu dài. Những kết quả trước mắt trong phần này tuy không phải là quá tốt nhưng cũng không phải là quá tệ. Phần kết

quả sẽ được trình bày ở phần sau cùng với việc so sánh giữa các thuật toán. Dưới đây là lược đồ thuật toán:



Hình 24 : Lược đồ thuật toán EBCR

Tuy nhiên để xây dựng một hệ dự đoán thành công, có rất nhiều phương pháp khác. Dưới đây là một trong những phương pháp mà áp dụng khá rõ bài toán kinh điển trong phân tích kỹ thuật ở lĩnh vực đầu tư chứng khoán. Sử dụng công cụ cực kì hữu hiệu là số Fibonacci trong lý thuyết sóng Elliott.

3.2.3. Mô hình ứng dụng Fibonacci trong lý thuyết sóng Elliott

Chúng ta đã thấy tầm quan trọng của con số Fibonacci và lý thuyết sóng trong lĩnh vực phân tích kỹ thuật (xem phụ lục) qua hàng chục năm ở các nước phát có nền kinh tế và thị trường chứng khoán phát triển đi trước. Trong lý thuyết sóng Elliot nó sử dụng chuỗi Fibonacci để dự đoán theo thời gian sự giao động của giá chứng khoán

trên thị trường. Do đó việc áp dụng mô hình dự đoán chuỗi thời gian áp dụng lý thuyết này một cách có phương pháp sẽ đưa đến cho chúng ta một mô hình dự đoán chứng khoán hiệu quả.

Trong phần này, dựa trên ý tưởng của những phương pháp tương tự đi trước, kế thừa ý tưởng sử dụng Fuzzy Logic và phương pháp mờ hóa của họ [29],[47] . Tuy nhiên trong quá trình tìm hiểu thì những kết quả đạt được không cao, chưa đạt được một hệ thống dự đoán tốt. Vì vậy trên tinh thần đó, đưa ra một số cải tiến và thực sự khá bất ngờ khi đạt được một số hiệu quả nhất định tốt hơn nhiều so với những kết quả mà họ đã đạt được (tuy nhiên ở đây chỉ là ở mức tương đối vì dữ liệu huấn luyện và đánh giá là khác nhau).

3.2.3.1. Chuỗi thời gian mờ (Fuzzy time-series) và lý thuyết sóng Elliot

Trước khi đề cập đến lý vào vấn đề chính xin nhắc lại một số vấn đề cơ bản liên quan, nó là nền tảng lý giải về phương pháp sẽ được trình bày ở phần sau.

Mô hình chuỗi thời gian mờ (Fuzzy time-series model)

Lý thuyết mờ đã được áp dụng từ lâu vào giải quyết những bài toán mang tính ngữ nghĩa. Tuy nhiên, theo tìm hiểu thì ban đầu những hệ thống “Mô hình chuỗi thời gian mờ” (Fuzzy time-series model) đã gặp nhiều thất bại, phải đến khoảng những năm 1993 thì một số chuyên gia trong lĩnh vực này mới xây dựng thành công mô hình này. Cho đến giờ nó vẫn được xem là một khung nền tảng (framework) cho những ai muốn xây dựng và giải quyết những vấn đề liên quan đến “Fuzzy time-series model”. Framework gồm 6 bước: (1) Xác định Universe Of Disource (UOD), (2) Xác định tập mờ cho các quan sát, (3) Mờ hóa quan sát, (4) Thiết lập mối quan hệ giữa các quan sát, (5) Dự đoán, (6) Giải mờ kết quả dự đoán. Dựa trên framework đã được xây dựng nhiều người đã đưa ra những mô hình của mình cho dự đoán chỉ các chỉ số chứng khoán khác nhau, nhưng kết quả đạt được so với nhiều phương pháp dự đoán khác như mạng noron ... thì có thể nói là chưa thực sự cải thiện nhiều.

Chuỗi Fibonacci và sóng Elliott

Chuỗi Fibonacci xuất hiện từ những năm 1202 do Leonardo Fibonacci đề xuất. Nhưng cho đến ngày nay nó vẫn là một công cụ rất mạnh và được sử dụng khác nhiều lĩnh vực khác nhau, trong dự đoán tài chính kinh tế là một ví dụ. Xin nhắc lại chuỗi Fibonacci như sau:

$$F_n = \begin{cases} 0 & n = 0 \\ 1 & n = 1 \\ F_{n-1} + F_{n-2}, & n > 1 \end{cases}$$

Trong đó theo phương pháp đẳng thức hồi tuyến tính bậc 2. Bằng phương pháp tính toán ta có thể biểu diễn biểu diễn chuỗi Fibonacci theo một cách khác như sau:

$$x(n) = Ax(n-1) + Bx(n-2)$$

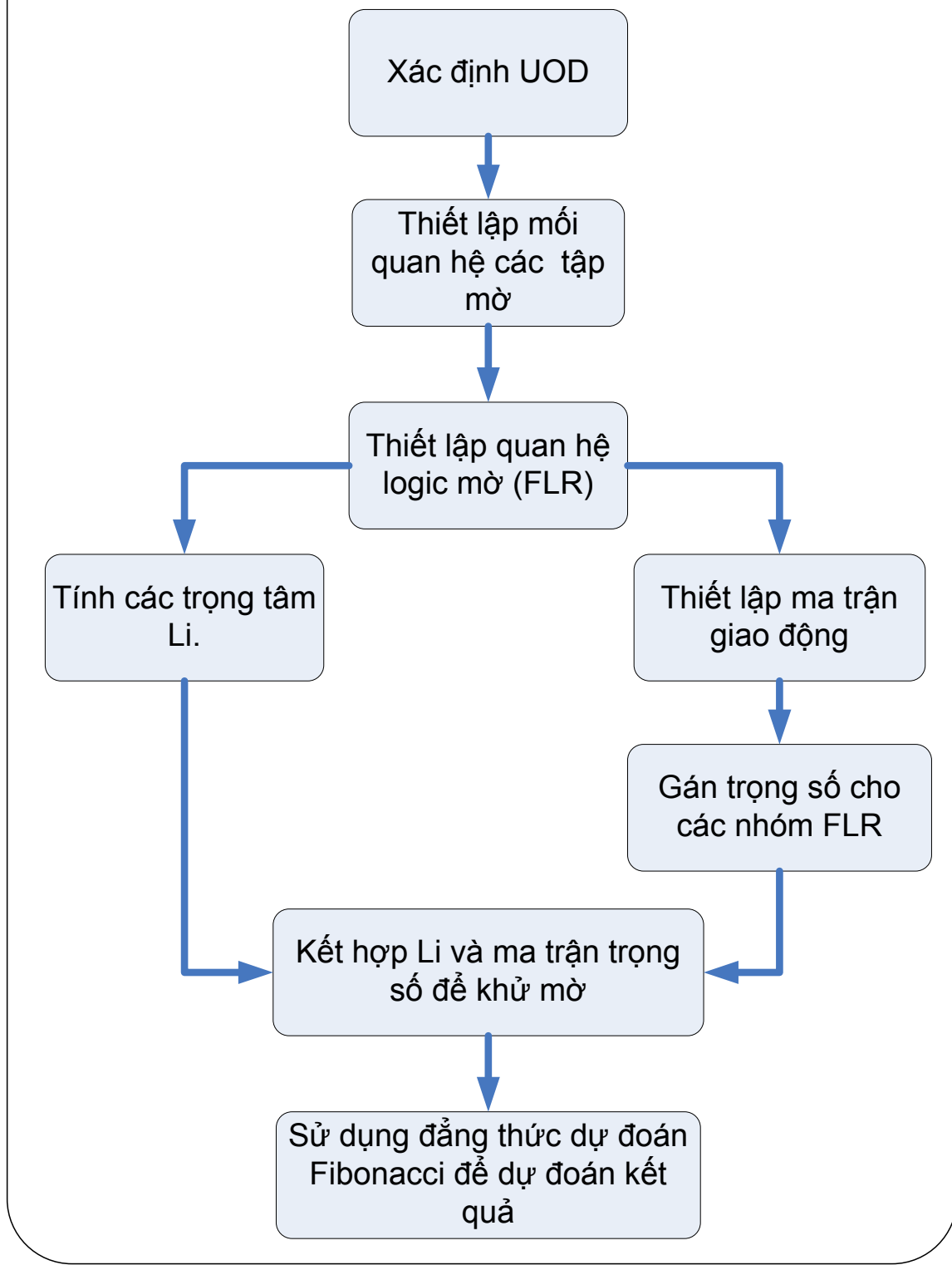
Với A và B là các hằng số. Công thức này sẽ được sử dụng trong quá trình xây dựng và ước lượng trong mô hình dự đoán ở phần sau.

Trong thị trường chứng khoán, các số Fibonacci nói chung được sử dụng trong phân tích kỹ thuật cùng với lý thuyết phân tích sóng Elliot để xác định mức hỗ trợ (support), mức kháng cự (resistance), và các đối tượng giá chứng khoán. Lý thuyết sóng Elliott cho rằng giá chứng khoán không chuyển động hỗn loạn như người ta nghĩ mà thị trường chuyển động theo một chu kỳ lặp lại, mà chính bản thân nó thể hiện tất cả mọi thứ của thị trường như: các hành động, cảm xúc và cả tâm lý của nhà đầu tư. Chính các mẫu lặp lại có thể chia nhỏ thành các “sóng”. Phần lý thuyết sóng sẽ không trình bày ở đây, có thể xem thêm ở phần phụ lục. Phần dưới đây sẽ đi vào trình bày chi tiết phương pháp xây dựng thuật toán.

3.2.3.2. Mô hình chuỗi thời gian mờ (Fuzzy time-series) chuẩn

Mô hình này đã được đề xuất từ lâu và khi bạn làm việc với “Mô hình thời gian mờ” mô hình dưới đây dựa trên framework có một số những cải tiến:

Mô hình chuỗi thời gian mờ (Fuzzy time-series model)



3.2.3.3. Mô hình đề xuất

Từ những lý thuyết ban đầu dựa trên mô hình nền tảng mô hình như trên, quá trình tìm hiểu và khám phá ra một số vấn đề như: xác định UOD hợp lý, phân phối dữ liệu thành những giá trị ngữ nghĩa phù hợp, và áp dụng lý thuyết phân tích chứng khoán trong mô hình chuỗi thời gian mờ. Để giải quyết những vấn đề được nêu ra xin đề xuất từ mô hình cơ bản sang một mô hình mới sử dụng Fibonacci vào dự đoán chứng khoán. Trong đó phần cải tiến nằm ở 3 phần sau:

- Sử dụng phương pháp đánh trọng mức độ giao động (fluctuation-weighted) mới với những giao động giá trong quá khứ.
- Sử dụng phương thức xác định trong tâm của mỗi giá trị được giải mờ (defuzzified).
- Áp dụng chuỗi Fibonacci $F_n = F_{n-1} + F_{n-2}$ vào quá trình dự đoán.

Điều thứ nhất, chúng ta tin rằng phương thức đánh trọng hiệu quả và rõ ràng sẽ giúp khảo sát, khai thác được chính xác những mối quan hệ trong chuỗi mờ và nhằm cải tiến mức độ chính xác trong mô hình truyền thống.

Thứ hai, theo một số tìm hiểu thì hầu hết các mô hình truyền thống hầu như đều sử dụng điểm “điểm giữa” (midpoint) trong trong khoảng ngữ nghĩa (linguistic interval) được sử dụng để khử mờ (defuzzified) nhằm tạo ra kết quả dự đoán. Đây là phương pháp dễ tính toán và được nhiều người công nhận. Tuy nhiên, khi bắt tay vào những bước đầu xây dựng ứng dụng và cài đặt thì nhận ra phương pháp này có nhiều khuyết điểm: (1) nó bỏ qua tính phân phối của dữ liệu của các quan sát cho mỗi khoảng ngữ nghĩa (linguistic interval) và điều này tất nhiên sẽ đi kèm với kết quả trả về với hệ số lỗi cao nếu trong trường hợp phân phối dữ liệu của các quan sát là không đối xứng. Do đó việc xác định trong tâm của quan sát như các giá trị được khử mờ sẽ giảm được mức độ lỗi nào đó trong một số trường hợp.

Thứ ba, áp dụng lý thuyết phân tích chứng khoán ở đây là áp dụng phương pháp phân tích kỹ thuật tất nhiên theo cảm quan nhận định thì nó sẽ cải tiến được mức độ chính xác cho mô hình.

Những vấn đề trên, tuy không phải là quá khó và khi đọc vào thì ai cũng có thể nghĩ hiểu và thấy nó thật dễ dàng. Nhưng thực sự để đi được đến những phân tích này phải mất khá nhiều thời gian và công sức từ lúc bắt đầu tìm hiểu và đến lúc xây dựng thành công ứng dụng thật sự với những cải tiến này thì kết quả đạt được một số kết quả nhất định.

Trình bày thuật toán

Thuật toán sẽ được trình bày chi tiết và đơn giản hóa trên một số ví dụ minh họa:

Bước 1 Xác định UOD một các hợp lý và chia bao nhiêu khoảng (intervals). Gồm 2 phần:

- (1) Ví dụ, giả sử cổ phiếu REE trong khoảng 7 năm hoạt động trên sàn giao dịch, giá cổ phiếu tập huấn luyện REE từ giá thấp nhất 10.2 (11/04/2003) và giá cao nhất 396 (30/05/2007). Xây dựng $UOD = [10, 400]$ đảm bảo toàn bộ giá trị huấn luyện đều nằm trong khoảng xác định này.
- (2) Chia UOD thành 10 phần và tính giá trị trung bình cho mỗi phần. Nếu số quan sát trong một phần xác định l_k lớn hơn giá trị trung bình của tất cả các phần thì phần đó nên được chia thành 2 phần nhỏ hơn.

Ví dụ. Sử dụng cổ phiếu REE ta thực hiện phép chia như sau:

Các khoảng phân chia	Miền giá trị
l_1	[10,48]

l_2	[49,87]
l_3	[88,126]
l_4	[127,165]
l_5	[166,204]
l_6	[205,243]
l_7	[244,282]
l_8	[283,321]
l_9	[322,360]
l_{10}	[361,400]

Bước 2 Thiết lập các tập mờ liên quan cho mỗi quan sát trong tập huấn luyện. Trong bước này. Các tập mờ L_1, L_2, \dots, L_k với các UOD được định nghĩa như sau:

$$\begin{aligned}
L_1 &= \frac{a_{11}}{u_1} + \frac{a_{12}}{u_2} + \frac{a_{13}}{u_3} + \dots + \frac{a_{1m}}{u_m} \\
L_2 &= \frac{a_{21}}{u_1} + \frac{a_{22}}{u_2} + \frac{a_{23}}{u_3} + \dots + \frac{a_{2m}}{u_m} \\
&\dots \\
L_k &= \frac{a_{k1}}{u_1} + \frac{a_{k2}}{u_2} + \frac{a_{k3}}{u_3} + \dots + \frac{a_{km}}{u_m}
\end{aligned}$$

Áp dụng vào bài toán, các giá trị a_{ij} thể mức độ phục thuộc của u_j vào tập mờ L_j .

Trong đó $a_{ij} \in [0,1]$. Trong bài toán này các giá trị a_{ij} xác định như sau:

$$\begin{aligned} L_1 &= \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \dots + \frac{0}{u_m} \\ L_2 &= \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \dots + \frac{0}{u_m} \\ &\dots \\ L_k &= \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \dots + \frac{1}{u_m} \end{aligned}$$

Từ tập các công thức trên, nếu giá trị hàm thành viên của giá ở L_j là lớn nhất thì hiển nhiên nó sẽ thuộc vào L_j . Mỗi quan sát sẽ được phân loại vào 10 tập, nếu giá trị nó ra ngoài tất cả các khoảng giá trị thì nó sẽ được gán vào 2 tập biên tùy vào giá trị của nó.

Dưới đây là bảng ví dụ:

Giá chứng khoán	Giá trị biến ngôn ngữ
50	L_2
103	L_3
40	L_1
60	L_2

200	L_5
340	L_9
150	L_4

Bước 3 Thiết lập FLR (Fuzzy Logical Relationship) cho chuỗi thời gian. Trong FLR được hình thành từ hai biến liên tục. Ví dụ $FLR(L_i, L_j)$ được thiết lập bởi $L_i(t)$ và $L_j(t-1)$. Bảng quan hệ dưới đây thể hiện một số quan hệ:

Bảng FLR
$L_1(t=1) \rightarrow L_1(t=2)$
$L_1(t=2) \rightarrow L_1(t=3)$
$L_2(t=3) \rightarrow L_1(t=4)$
$L_1(t=4) \rightarrow L_1(t=5)$
$L_2(t=5) \rightarrow L_3(t=6)$
$L_3(t=6) \rightarrow L_3(t=7)$
$L_3(t=7) \rightarrow L_4(t=8)$

Bước 4 Dựa trên mỗi quan hệ FLR ta tạo ra bảng ma trận dịch chuyển của chúng như sau. Mỗi tập mờ L_i và L_j có một quan hệ trong bảng quan hệ FLR thì sẽ được cộng +1 vào bảng tương ứng dưới đây:

$P(t)$	$P(t+1)$									
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}
L_1	1	2	0	0	0	0	0	0	0	0
L_2	1	0	1	0	0	0	0	0	0	0
L_3	0	0	1	1	0	0	0	0	0	0
L_4	0	0	0	1	3	0	0	0	0	0
L_5	0	0	0	0	0	1	0	0	0	0
L_6	0	0	0	0	1	0	0	0	0	0
L_7	0	0	0	0	0	2	0	0	0	0
L_8	0	0	0	0	0	0	1	0	0	0
L_9	0	0	0	0	0	0	0	0	2	0
L_{10}	0	0	0	0	0	0	0	0	1	1

Bước 5 Mỗi nhóm FLR sẽ được gán một trọng số dựa vào bảng ma trận dịch chuyển. Mỗi giả sử nếu có quan hệ $FLR(L_i, L_j)$ thì +1 vào mỗi quan hệ. Công thức tính trọng số đơn giản thực hiện như sau. Trong đó k là số lần xảy ra quan hệ $FLR(L_i, L_j)$

$$W_{L_i \rightarrow L_j} = \sum k$$

Sử dụng công thức ta tính được bảng sau:

$P(t)$	$P(t + 1)$										
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	$\sum_{k=0}^9 W_k$
L_1	1	2	0	0	0	0	0	0	0	0	3
L_2	1	0	1	0	0	0	0	0	0	0	2
L_3	0	0	1	1	0	0	0	0	0	0	2
L_4	0	0	0	1	3	0	0	0	0	0	4
L_5	0	0	0	0	0	1	0	0	0	0	1
L_6	0	0	0	0	1	0	0	0	0	0	1
L_7	0	0	0	0	0	2	0	0	0	0	2
L_8	0	0	0	0	0	0	1	0	0	0	1
L_9	0	0	0	0	0	0	0	0	2	0	2
L_{10}	0	0	0	0	0	0	0	0	1	1	2

Sau khi tính tổng trọng số của mỗi tập mờ, ta thực hiện chuẩn hóa các giá trị về khoảng $[0,1]$. Đơn giản bằng cách chia cho tổng trọng số của mỗi dòng bằng công thức sau:

$$W_n(t) = [W'_1, W'_2, \dots, W'_9] = \left[\frac{W_1}{\sum_{k=0}^9 W_k}, \frac{W_2}{\sum_{k=0}^9 W_k}, \dots, \frac{W_9}{\sum_{k=0}^9 W_k} \right]$$

Ta được bảng kết quả sau:

$P(t)$	$P(t+1)$										
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	$\sum_{k=0}^9 w_k$
L_1	1/3	2/3	0	0	0	0	0	0	0	0	3
L_2	1/2	0	1/2	0	0	0	0	0	0	0	2
L_3	0	0	1/2	1/2	0	0	0	0	0	0	2
L_4	0	0	0	1/4	3/4	0	0	0	0	0	4
L_5	0	0	0	0	0	1	0	0	0	0	1
L_6	0	0	0	0	1	0	0	0	0	0	1
L_7	0	0	0	0	0	1	0	0	0	0	2
L_8	0	0	0	0	0	0	1	0	0	0	1
L_9	0	0	0	0	0	0	0	0	1	0	2
L_{10}	0	0	0	0	0	0	0	0	1/2	1/2	2

Bước 6 Tính toán giá trị trọng tâm v_i của mỗi khoảng (linguistic interval) L_i sử dụng công thức:

$$v_i = \frac{\sum_{k=1}^n u_{ik} x_k}{\sum_{k=1}^n u_{ik}}$$

Xác định trọng tâm của các khoảng (interval) ta được ma trận trọng tâm $L_{df}(t)$ là ma trận kích thước $1 \times N$ với N là số khoảng (interval).

Bước 7 Thực hiện khử mờ, từ hai bước 5 và 6 ta được hai ma trận (1) ma trận trọng số được chuẩn hóa (2) ma trận trọng tâm. Dựa trên hai ma trận này ta suy ra được ma trận dự đoán khởi tạo:

$$Forecast(t+1) = L_{df}(t) * W_n(t)$$

Bước 8 Ở bước này là ta sử dụng đẳng thức Fibonacci nhằm tạo ra kết quả dự đoán liên tục. Bằng cách sử dụng 2 biến tuyến tính α và β , mô hình đề xuất có thể điều chỉnh sao cho khả năng đạt được dự đoán là nhỏ nhất có thể trong tập huấn luyện, như vậy sẽ tạo ra kết quả dự đoán đáng tin cậy:

$$Fibonacci(t+1) = P(t) + \alpha(Forecast(t+1) - P(t)) + \beta(Forecast(t) - P(t-1))$$

Trong bước này, hai giá trị α và β nhận hai giá trị trong khoảng $[-1,1]$. Các biến được xác định khi thực hiện dự đoán mà đạt hiệu quả tốt nhất trong tập huấn luyện. Tuy nhiên có một vấn đề là trong Fibonacci thì hai biến này là hằng số, nhưng trong thực tế thì hai biến này luôn được biến đổi phù hợp với ngữ cảnh và tập huấn luyện. Và trong thực tế điều này là có thể chấp nhận được.

Chương 4. KẾT QUẢ VÀ ĐÁNH GIÁ

4.1 Phương pháp CBR cho đánh giá chứng khoán

4.1.1. Dữ liệu đánh giá

Vấn đề này phức tạp nhất khi thực hiện xây dựng cũng như đánh giá mô hình. Toàn bộ dữ liệu sẽ không phải do bất cứ một chuyên gia nào trong lĩnh vực chứng khoán xây dựng. Mà dựa ý kiến của nhiều người, những người mà có một số kinh nghiệm được đào tạo bài bản về các chiến lược chứng khoán, đồng thời cũng đã tham gia đầu tư trong thị trường một khoảng thời gian. Những nhận định được thu lại và thanh lọc để có một bộ dữ liệu hoàn chỉnh. Theo bản thân đánh giá thì vấn đề tạo dữ liệu ở đây chưa được thuyết phục. Xong, vì nhiều lý do nên chỉ có thể thực hiện phương pháp này. Tuy nhiên, điều này không phải là không tốt, vì dữ liệu được thu thập trên một bình diện nhiều nhà đầu tư, điều này làm cho bộ cơ sở tình huống trở nên khách quan hơn so với chỉ thu thập từ một chuyên gia nào đó trong lĩnh vực. Ngoài ra, nhiều người dù sao cũng có cái nhìn bao quát và rộng hơn trong tất cả các tình huống.

Bộ dữ liệu bao gồm 300 tình huống ban đầu, với mức độ phân phối đồng đều các trường hợp. Ngoài ra sẽ có 50 tình huống khác được sử dụng để kiểm tra và đánh giá mô hình. 50 tình huống này hoàn toàn không nằm trong bộ cơ sở tình huống ban đầu. Điều này nhằm tăng tính khách quan trong quá trình kiểm tra những kết quả đạt được. Do số lượng tình huống khá lớn nên sẽ không trình bày trong khóa luận. Toàn bộ sẽ được nằm trong bộ cơ sở dữ liệu của hệ thống CBR.

4.1.2. Một số kết quả đạt được

Dựa trên 50 tình huống riêng khác dựa vào đánh giá đạt được một số kết quả sau:

Số lượng tình huống trong cơ sở tình huống	Số lượng tình huống đưa vào đánh giá	Số lượng tình huống đúng	Phần trăm kết quả đạt được
300	50	44	88%

Bảng 11 : Bảng kết quả đạt được trong mô hình CBR đánh giá cổ phiếu

Một số ví dụ khi thực hiện đánh giá tình huống trong số 50 mẫu :

Tình huống 1:

<i>Các thuộc tính</i>	<i>Giá trị lựa chọn</i>	<i>Trọng số</i>
<i>EPS</i>	<i>15%</i>	<i>0.57</i>
<i>P/E</i>	<i>18</i>	<i>0.81</i>
<i>P/B</i>	<i>2.5</i>	<i>0.5</i>
<i>ROE</i>	<i>25%</i>	<i>0.79</i>
<i>So với P/E ngành</i>	<i>Thấp hơn</i>	<i>0.56</i>
<i>So với P/B ngành</i>	<i>Thấp hơn</i>	<i>0.39</i>
<i>Lợi nhuận</i>	<i>18%</i>	<i>0.68</i>
<i>Lợi nhuận cùng kì</i>	<i>13%</i>	<i>0.36</i>
<i>Khả năng cạnh tranh</i>	<i>Cạnh tranh</i>	<i>0.37</i>
<i>Ưu đãi</i>	<i>Có</i>	<i>0.68</i>
<i>Xu hướng tăng giá</i>	<i>Không</i>	<i>0.7</i>
<i>Thanh khoản</i>	<i>Không</i>	<i>0.45</i>
<i>Kết quả ban đầu</i>	<i>Cổ phiếu tốt</i>	
<i>Kết quả đánh giá</i>	<i>Cổ phiếu tốt</i>	

Tình huống 2:

<i>Các thuộc tính</i>	<i>Giá trị lựa chọn</i>	<i>Trọng số</i>
-----------------------	-------------------------	-----------------

<i>EPS</i>	<i>7.4%</i>	<i>0.5</i>
<i>P/E</i>	<i>27.4</i>	<i>0.79</i>
<i>P/B</i>	<i>4</i>	<i>0.5</i>
<i>ROE</i>	<i>19.23%</i>	<i>0.76</i>
<i>So với P/E ngành</i>	<i>Tương đương</i>	<i>0.64</i>
<i>So với P/B ngành</i>	<i>Tương đương</i>	<i>0.58</i>
<i>Lợi nhuận</i>	<i>8.7%</i>	<i>0.61</i>
<i>Lợi nhuận cùng kì</i>	<i>10.5%</i>	<i>0.39</i>
<i>Khả năng cạnh tranh</i>	<i>Cạnh tranh</i>	<i>0.34</i>
<i>Ưu đãi</i>	<i>Không</i>	<i>0.4</i>
<i>Xu hướng tăng giá</i>	<i>Không</i>	<i>0.53</i>
<i>Thanh khoản</i>	<i>Không</i>	<i>0.59</i>
<i>Kết quả ban đầu</i>	<i>Cổ phiếu xấu</i>	
<i>Kết quả đánh giá</i>	<i>Cổ phiếu xấu</i>	

Tình huống 3:

<i>Các thuộc tính</i>	<i>Giá trị lựa chọn</i>	<i>Trọng số</i>
<i>EPS</i>	<i>18%</i>	<i>0.5</i>
<i>P/E</i>	<i>27.4</i>	<i>0.39</i>
<i>P/B</i>	<i>4</i>	<i>0.5</i>
<i>ROE</i>	<i>18.23%</i>	<i>0.55</i>
<i>So với P/E ngành</i>	<i>Tương đương</i>	<i>0.71</i>
<i>So với P/B ngành</i>	<i>Cao hơn</i>	<i>0.55</i>
<i>Lợi nhuận</i>	<i>18.7%</i>	<i>0.69</i>
<i>Lợi nhuận cùng kì</i>	<i>10.45%</i>	<i>0.45</i>
<i>Khả năng cạnh tranh</i>	<i>Cạnh tranh</i>	<i>0.39</i>

<i>Ưu đãi</i>	<i>Có</i>	<i>0.48</i>
<i>Xu hướng tăng giá</i>	<i>Không</i>	<i>0.47</i>
<i>Thanh khoản</i>	<i>Có</i>	<i>0.47</i>
<i>Kết quả ban đầu</i>	<i>Cổ phiếu thường</i>	
<i>Kết quả đánh giá</i>	<i>Cổ phiếu thường</i>	

Trên đây chỉ là một số tình huống truy vấn trong số 50 tình huống đưa vào đánh giá mô hình. Quá trình đánh giá dựa trên tiêu chuẩn “Một tình huống đánh giá đưa vào sẽ được đánh giá trước đó. Sau khi đưa vào mô hình của chương trình. Nếu kết quả đạt được cùng kết quả so với kết quả đã đánh giá trước đó thì xem như việc đánh giá tình huống thành công”. Dựa trên số lượng tình huống được xem là thành công để xác định được tỉ lệ phần trăm mà mô hình đánh giá đạt được.

4.1.3. Đánh giá

Mô hình hoạt động khá tốt dựa trên bộ mẫu học và bộ mẫu đánh giá. Kết quả đạt được tuy không phải là quá cao nhưng nó cũng có thể chấp nhận được. Tuy nhiên, những tình huống cũng còn mang nặng tính lý thuyết và cũng chưa có điều kiện thực hiện những đánh giá với các tình huống thực tế, đây là một yêu cầu khó và cũng cần một khoảng thời gian khá dài nhằm đánh giá được hiệu quả của nó. Nhưng ở đây dựa trên những gì đạt được cũng có những nhìn nhận về mô hình như sau:

4.1.3.1. Ưu điểm

- Mô hình phân loại khá chính xác các tình huống dựa trên những bộ tình huống kiểm tra ở trên.
- Khá linh động trong vấn đề lựa chọn các chỉ tiêu tình huống khác nhau dựa trên bộ trọng số do người sử dụng nhập vào.

4.1.3.2. Nhược điểm

- Tuy nhiên, vấn đề đánh trọng cũng lại có một yếu điểm khác. Yêu cầu người sử dụng phải xác định được chính xác một số trọng chính để thực hiện đánh chỉ số tình huống. Điều này thực sự cũng là không phải đơn giản.
- Một vấn đề khác mà mô hình chưa giải quyết là còn khá phức tạp và đòi hỏi những người sử dụng phải có một số tìm hiểu đơn giản về các kiến thức liên quan. Tuy nhiên, ở đây các thông số này đã được công bố trên một số website [11],[12]. Vì vậy người sử dụng chỉ đơn giản tìm các công bố, các cáo bạch hàng tháng, hàng quý của công ty phát hành cổ phiếu đó và đưa vào ứng dụng.

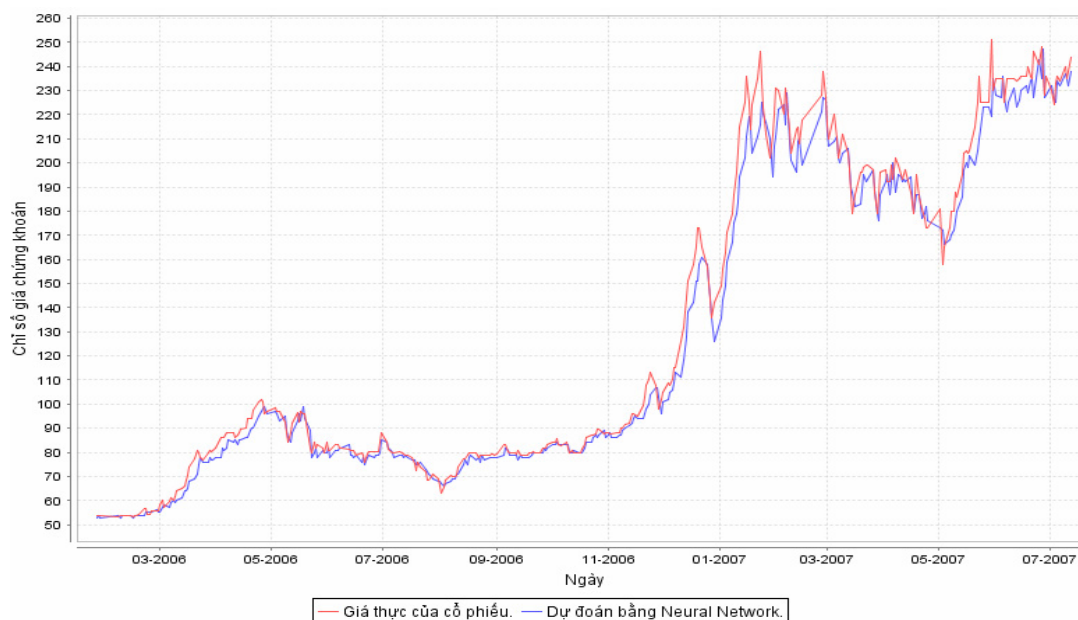
4.1.4. Đề xuất phương pháp cải tiến trong tương lai

Bài toán hiện tại vẫn chưa được giải quyết tốt. Vẫn còn khá nhiều vướng mắc chưa giải quyết được như trình bày ở trên. Việc đánh trọng như đề xuất ban đầu chỉ phụ thuộc nhiều vào những thuộc tính có tỉ số đánh trọng cao. Điều này xảy ra trường hợp quá khớp trong việc đối sánh các tình huống sử dụng các độ đo như Manhattan hoặc Euclidean. Đối với bài toán dự đoán chứng khoán việc sử dụng một vài tiêu chí chủ đạo thực sự chưa phải là một phương sách hiệu quả, vì trên các thuộc tính đánh giá đều có những mối quan hệ nhất định và có một tầm quan trọng đặc biệt trong mỗi tình huống, vì vậy thực sự là khó để người dùng có thể thấy được những mối quan hệ giữa chúng để có thể có những phương thức chọn lựa trọng số hiệu quả. Do đó, một trong những đề xuất cho bài toán này là dựa vào tập dữ liệu học có thể xác định được càng chính xác những mối quan hệ tương hỗ giữa chúng càng tốt. Do vậy khi thực hiện đánh trọng thì những trọng số được hình thành một mối quan hệ ngầm hay nói đúng hơn thì nếu trọng số giữa chúng sẽ có những mối quan hệ ràng buộc nhau theo một tiêu chí nào đó.

4.2 Các phương pháp đạt được trong dự đoán kĩ thuật

Sử dụng 3 mô hình khác nhau cho dự đoán chứng khoán, mỗi phương pháp đạt được những kết quả khác nhau mỗi phương pháp có những ưu điểm cũng như những khuyết

điểm riêng. Kết quả đạt được là khá khả quan trong lĩnh vực phân tích kỹ thuật. Dưới đây là một số những biểu đồ thể hiện kết quả đạt được.



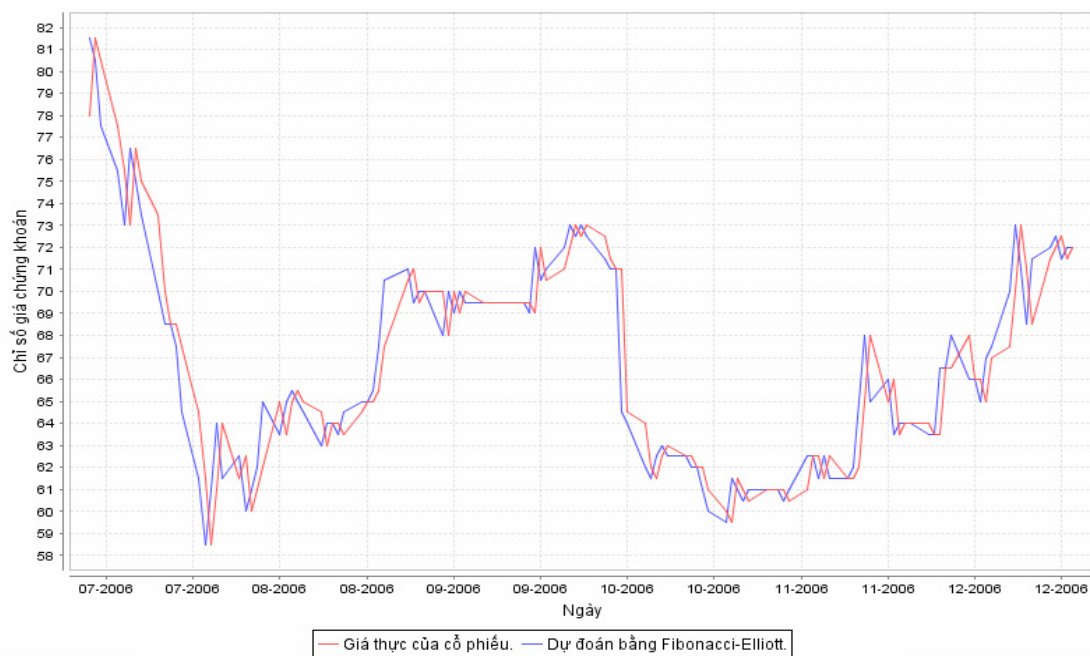
Hình 26 : Dự đoán giá KDC sử dụng mạng nơron (25/01/2006 - 12/07/2007)



Hình 27 : Dự đoán giá cổ phiếu BBC bằng mạng nơron (01/01/2005 - 12/07/2007)



Hình 28 : Dự đoán cổ phiếu BMC bằng E-CBR (01/06/2007 - 01/07/2007)



Hình 29 : Dự đoán giá cổ phiếu STB bằng phương pháp Fibonacci (12/07/2006 - 12/07/2007)



Hình 30 : So sánh giữa mô hình RW và Fibonacci

Đây là một số kết quả nhỏ trong quá trình kiểm tra và dự đoán kết quả trong một số khoảng thời gian trong quá khứ. Phần dưới là những kết quả chi tiết cho mỗi bài toán. Chúng ta có thể so sánh kết quả đạt được của từng thuật toán khác nhau. Những kết quả này được thực hiện kiểm định trên một số khoảng thời gian khác nhau. Hệ số lỗi được tính dựa trên 3 phương pháp tính hệ số lỗi khác nhau.

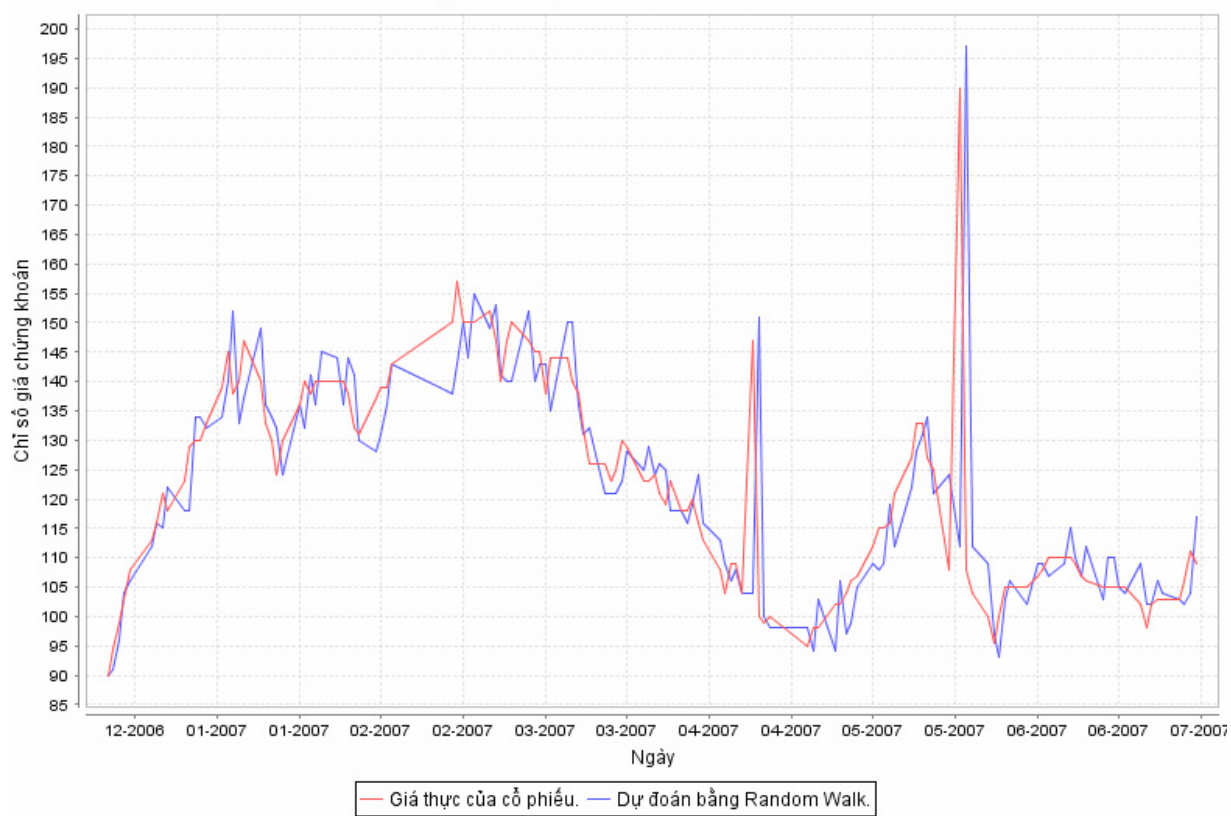
Ta sẽ thực hiện việc so sánh kết quả của các phương pháp với một số những cổ phiếu đại diện cho các nhóm ngành khác nhau. Theo trang web www.vietstock.com.vn thì phân chia thành 8 nhóm ngành khác nhau. Ta sẽ chọn một nhóm ngành gồm 2 hoặc 4 cổ phiếu đại diện để thực hiện việc so sánh giữa các phương pháp.

Tên nhóm ngành	Số doanh nghiệp	Cổ phiếu đại diện
Nguyên liệu cơ bản	4	BMC, SSC
Hàng tiêu dùng	31	ABT, GMC, VNM, SGC
Tài chính	3	STB, VF1
Y tế, sức khỏe	3	DMC, IMP
Hàng công nghiệp	37	BTC, LBM, REE, VTA
Dịch vụ	23	GMD, PGC, TNA
Kỹ thuật, công nghệ	3	FPT, SAM
Hàng thiết yếu	5	RHC, VSH

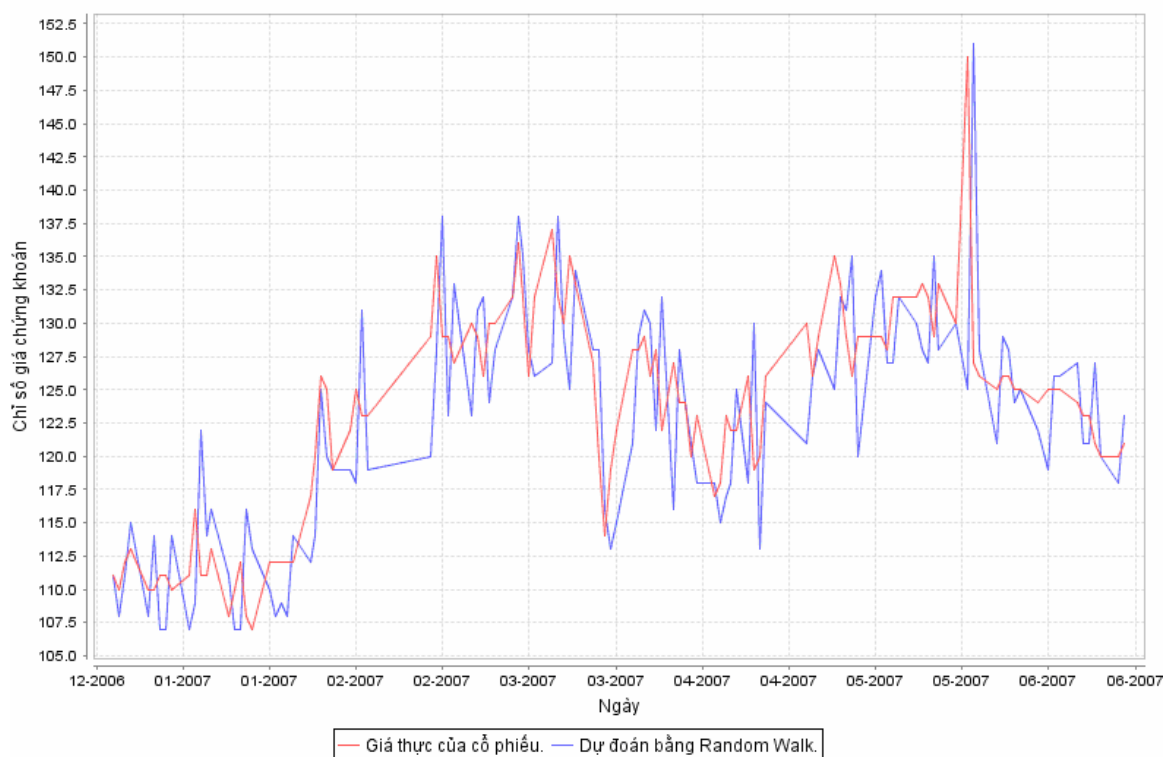
Bảng 12 : Cổ phiếu đại diện cho mỗi nhóm ngành được niêm yết trên thị trường cho việc so sánh các phương pháp

4.2.1. Phương pháp Random Walk

Tất nhiên là phương pháp đạt hệ số lỗi cao nhất, việc sử dụng phương pháp này để chứng minh rằng việc sử dụng chuỗi ngẫu sinh ngẫu nhiên trong quá trình dự đoán chứng khoán là một điều không thể vì hệ số lỗi đạt được của phương pháp này là khá lớn. Mặc dầu đối với thị trường Việt Nam sự giao động giá trong một ngày bị giới hạn $[-5\%, +5\%]$. Nhưng những kết quả đạt được so với sự biến động giá vẫn khá lớn trên dưới 3%. Dưới đây là một số kết quả đạt được:



Hình 31 : Random Walk trên cổ phiếu ABT (01/01/2006 - 12-07-2007)



Hình 32 : Biểu đồ RW cổ phiếu RAL (12/2006 - 6/2007)

Dưới đây là một số cổ phiếu đại diện được sử dụng để thực hiện quá trình kiểm chứng trên mô hình Random Walk. Kết quả này có thể bị thay đổi sau mỗi lần dự đoán vì ở đây ta thực hiện sinh quá trình sinh ngẫu nhiên.

Tên cổ phiếu	Khoảng thời gian	RMSE	MAPE	MAE
ABT	13/07/2007-19/07/2007	12.47	4.68	5.9
BMC	13/07/2007-19/07/2007	83.32	7.35	29.7
BTC	13/07/2007-19/07/2007	1.02	3.94	0.69
DMC	13/07/2007-19/07/2007	6.13	3.4	4.27
FPT	13/07/2007-19/07/2007	32.41	4.2	20.5
GMC	13/07/2007-19/07/2007	3.07	3.8	2.3

GMD	13/07/2007-19/07/2007	10.78	4.44	5.02
IMP	13/07/2007-19/07/2007	7.044	3.79	4.68
LBM	13/07/2007-19/07/2007	2.37	4.734	1.725
PGC	13/07/2007-19/07/2007	3.77	4.34	2.8
REE	13/07/2007-19/07/2007	2.367	4.18	2.675
RHC	13/07/2007-19/07/2007	2.29	3.3	1.8
SAM	13/07/2007-19/07/2007	20.9	5.23	6.44
SGC	13/07/2007-19/07/2007	2.17	3.42	1.71
SSC	13/07/2007-19/07/2007	6.95	4.23	2.93
STB	13/07/2007-19/07/2007	7.73	3.87	3.886
TNA	13/07/2007-19/07/2007	1.96	3.436	1.51
VF1	13/07/2007-19/07/2007	2.191	3.838	1.215
VNM	13/07/2007-19/07/2007	10.33	3.72	6.89
VSH	13/07/2007-19/07/2007	4.02	4.0	2.9
VTa	13/07/2007-19/07/2007	2.07	4.72	1.205

Bảng 13 : Bảng kết quả dự đoán bằng RW trên một số cổ phiếu đại diện

Trong 3 hệ số lỗi hệ số lỗi dễ cảm nhận nhất là MAPE, các chỉ số lỗi này hầu như không thay đổi nhiều các giá trị đều > 3%. Như đã trình bày ở trên.

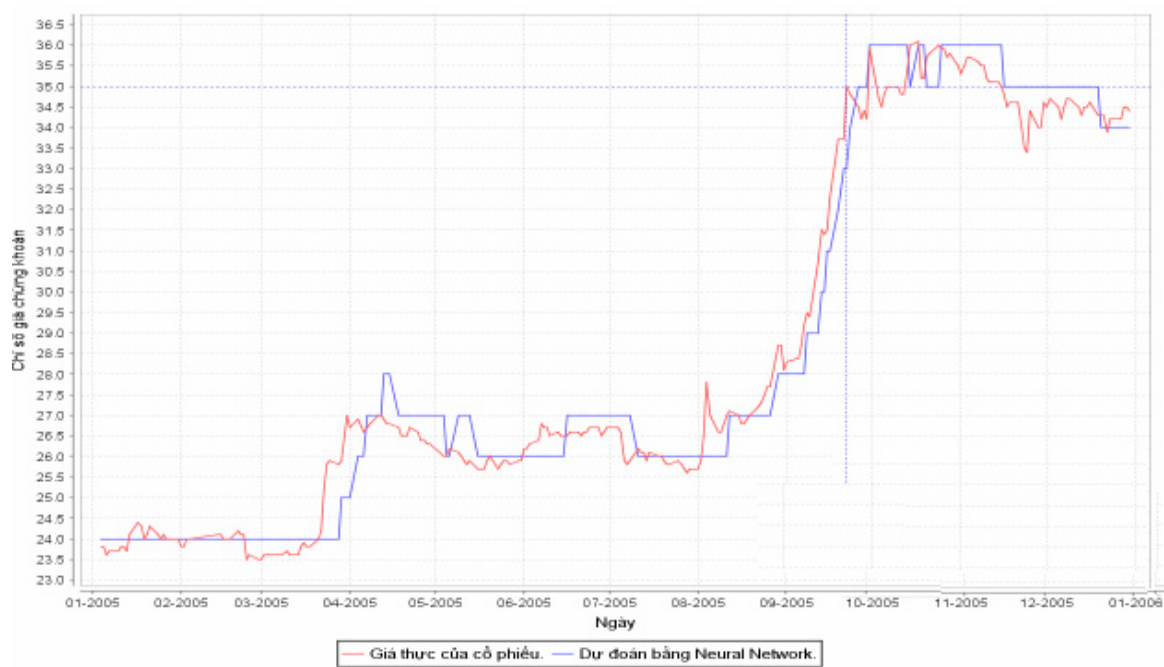
4.2.2. Phương pháp mạng nơron

4.2.2.1. Kết quả đạt được

Phương pháp này trong một số trường hợp đạt được kết quả khá tốt, nhưng trong một số trường hợp thì cũng chỉ ở mức độ chấp nhận được. Ngoài ra phương pháp có một khuyết điểm là yêu cầu một khối lượng cơ sở dữ liệu khá lớn trong quá trình thực hiện việc học trong ứng dụng. Dưới đây là một số hình ảnh thể hiện kết quả đạt được.



Hình 33 : Dự đoán REE bằng mạng nơron (3/2006 – 4/2007)



Hình 34 : Dự đoán cổ phiếu REE bằng mạng nơron (1/2005 – 1/2006)

Dưới đây là một số những kết quả đạt được khi thực hiện dự đoán bằng phương pháp mạng nơron:

Tên cổ phiếu	Khoảng thời gian	RMSE	MAPE	MAE
ABT	13/07/2007-19/07/2007	2.74	2.428	2.593
BMC	13/07/2007-19/07/2007	17	3.55	14.9
BTC	13/07/2007-19/07/2007	1.162	3.49	1.053
DMC	13/07/2007-19/07/2007	5.119	2.813	3.83
FPT	13/07/2007-19/07/2007	4.64	1.379	4.011
GMC	13/07/2007-19/07/2007	3.426	5.346	2.85
GMD	13/07/2007-19/07/2007	1.07	0.657	0.922
IMP	13/07/2007-19/07/2007	6.845	3.782	5.478

LBM	13/07/2007-19/07/2007	1.036	1.594	0.952
PGC	13/07/2007-19/07/2007	0.931	1.17	0.654
REE	13/07/2007-19/07/2007	2.236	1.331	1.961
RHC	13/07/2007-19/07/2007	0.93	1.657	0.759
SAM	13/07/2007-19/07/2007	1.949	1.083	1.682
SGC	13/07/2007-19/07/2007	0.521	0.847	0.379
SSC	13/07/2007-19/07/2007	1.248	1.443	1.013
STB	13/07/2007-19/07/2007	2.923	4.425	2.748
TNA	13/07/2007-19/07/2007	1.628	2.572	1.394
VF1	13/07/2007-19/07/2007	2.718	8.62	2.6
VNM	13/07/2007-19/07/2007	1.561	0.78	1.397
VSH	13/07/2007-19/07/2007	0.624	0.934	0.567
VT A	13/07/2007-19/07/2007	0.945	3.03	0.78

Bảng 14 : Bảng kết quả dự đoán bằng mạng nơron trên một số cổ phiếu đại diện

4.2.2.2. Đánh giá thuật toán

Như ta thấy thì thuật toán chỉ chạy tốt và đạt được hệ số lỗi đủ nhỏ trên một số những cổ phiếu, còn lại với một số những cổ phiếu khác thì kết quả chẳng hơn gì so với phương pháp RW. Có thể nhìn nhận thuật toán như sau.

Ưu điểm

- Với những cổ phiếu có lượng dữ liệu đủ lớn thì kết quả nhận dạng là rất tốt.
- Thuật toán chạy nhanh sau quá trình huấn luyện.

- Nếu có một tình huống tương tự xảy ra trong quá khứ thì khả năng dự đoán là rất cao.
- Thuật toán ít bị ảnh hưởng bởi dữ liệu nhiễu trong quá trình dự đoán hơn các thuật toán khác sẽ đề cập ở phần sau.
- Với những cổ phiếu có khối lượng dữ liệu > 7 năm thì phương pháp đạt được hệ số lỗi rất tốt và chấp nhận được.

Khuyết điểm

- Thuật toán đòi hỏi phải có lượng dữ liệu lớn, ít nhất là 7 năm để có thể thực hiện quá trình huấn luyện, và thời gian để thực hiện huấn luyện là khá lâu.
- Đối với những cổ phiếu mới niêm yết gần đây thì thuật toán này không thực sự hiệu quả. Như ta thấy hệ số lỗi đạt được khi thực hiện dự đoán trong bảng kết quả ở trên.
- Thuật toán gặp một vấn đề là xảy ra trường hợp quá khớp dữ liệu trong quá trình học.

4.2.2.3. Đề xuất phương pháp cải tiến trong tương lai

Theo đánh giá những gì đã đạt được, cái quan trọng nhất trong quá trình dự đoán bằng mạng nơ-ron là việc biến đổi dữ liệu, xác định mối quan hệ tương hỗ giữa những giá ở thời điểm trước và giá ở thời điểm sau. Do đó, một trong những định hướng tương lai khi phát triển kỹ thuật áp dụng mạng nơ-ron là tập trung vào việc tìm hiểu những mối quan hệ giữa chúng. Có thể sẽ có một mô hình biến đổi khác hợp lý hơn trong việc chuẩn hóa và biến đổi các giá trị của thuật toán. Hiện giờ vẫn chưa có ý tưởng nào khả dĩ hiệu quả. Tuy nhiên, trong tương lai hi vọng sau một quá trình tìm hiểu khác sẽ đưa ra những phương pháp hiệu quả hơn. .

4.2.3. Sử dụng các thông số kỹ thuật trong mạng nơron

4.2.3.1. Một số kết quả đạt được

Phương pháp này chỉ là một đề xuất cho một định hướng khác với mục đích áp dụng các thông số kỹ thuật trong phân tích kỹ thuật vào việc dự đoán với kì vọng mong muốn có một sự cải tiến mạnh mẽ trong bài toán dự đoán giá cả trong tương lai. Tuy nhiên, không đạt được kết quả như một số những phương pháp trước, phương pháp này chỉ mang tính đề xuất và kết quả thực sự lại không cao. Nhưng dù sao cũng là một phương pháp đáng chú ý và theo bản thân thì vẫn có những khả năng phát triển loại phương pháp này trong bài toán chứng khoán. Dưới đây sẽ là bảng kết quả một số các thông số kỹ thuật trên một vài cổ phiếu có khối lượng dữ liệu quá khứ lớn. Ở đây xin chọn 1 cổ phiếu đại diện là ABT để minh họa một số kết quả.

Ví dụ. Thông tin các thông số kỹ thuật của cổ phiếu ABT và dự đoán

Open	Close	Volume	AccDist	Rsi	Stock%D	Stock%K	%R	ArmsEM	Obv	Pvt	%Change	GTT	GDD
95.5	100	1514	92437	49.413	34.47919	7.905882	-94.7368	0.004458	130178	107.199	-0.24	105.0	72.0
100	105	3379	93713	43.069	28.91876	22.4	-89.4737	0	132281	75.7	-0.068	107.0	93.1
105	105	4534	-17214	50.976	27.397	36.47407	-89.4737	0.001764	32040	0	0.03	108.0	107
105	105	2053	-14138	51.176	27.6345	50.54815	-89.4737	0	31234	0	0.15	110.0	120
105	107	4645	-9493	51.367	31.76122	67.58519	-87.3684	4.31E-04	34744	39.10476	0.14	110.0	122
107	108	5377	-10694	52.579	37.93437	75.38519	-86.3158	0	54998	43.41121	0.14	110.0	123
108	110	4065	-6629	52.390	45.14869	80.18519	-84.6561	3.69E-04	51024	99.57407	0.14	109.0	120
110	110	4684	-7448	52.390	53.32435	86.11111	-84.6561	0	50897	0	0.13	107.0	125
110	110	1713	-7448	51.619	63.84074	92.03704	-84.6561	0	50205	0	0.13	106.0	124
110	109	1683	-7138	51.411	72.08241	88.33333	-85.7143	-8.91E-04	50486	-15.5727	0.12	105.0	123
109	107	501	-3801	50.604	76.68981	73.33333	-87.8307	0.001996	46648	-30.8807	0.12	105.0	122
107	106	408	3824	50	77.8713	60	-88.8889	-0.01225	39023	-4.68224	0.11	105.0	119
106	105	534	5846	49.595	74.83981	43.33333	-89.9471	9.36E-04	35933	-3.84906	0.10	105.0	117
105	105	1497	1648	48.553	68.83333	27.33333	-89.9471	6.68E-04	33232	0	0.11	105.0	117
106	105	42260	-40612	47.907	61.39352	20.66667	-89.9471	-1.18E-05	27261	0	0.12	102.0	117
105	105	26490	-36091	47.907	54.0463	27.33333	-89.9471	0	27595	0	0.12	98.0	117
105	105	7280	-31236	47.689	45.125	20.66667	-89.9471	0	22740	0	0.12	102.0	118
103	102	36030	-64640	45.974	36.25	17.33333	-93.1217	-6.94E-05	-15916	-208	0.13	103.0	119
98	98	12320	-81681	43.803	28.75	13.33333	-97.3545	-3.25E-04	-32957	-1412.94	0.13	103.0	115
98	102	5110	-76571	43.318	23.5119	18.09524	-93.1217	0.001174	-30140	502.8571	0.13	103.0	115
103	103	27970	-48601	43.562	21.30952	25.71429	-92.0635	8.94E-05	-422	50.09804	0.14	106.0	115
103	103	21170	-41425	44.713	23.60714	45.71429	-92.0635	0	120	0	0.14	111.0	117
103	103	305	-37566	45.11111	29.2381	65.71429	-92.0635	0	5315	0	0.14	109.0	117
103	106	2122	-42093	49.52607	36.53571	85.71429	-88.8889	0	11604	8.883495	0.028302	n/a	118
106	111	5050	-46440	20.52239	45.7381	94.28571	-83.5979	0	12307	100.0943	-0.01802	n/a	120

GTT : Là giá thực 4 ngày sau đó.

GDD : Là giá trị dự đoán 4 ngày sau đó.

Phương pháp này thực sự chưa được hiệu quả, có thể một phần là do kỹ thuật cài đặt, cũng có thể là có những kỹ thuật áp dụng vào chưa thật hợp lý. Hi vọng trong tương lai gần sẽ có những phương pháp hiệu quả hơn và đạt được những kết quả tốt hơn.

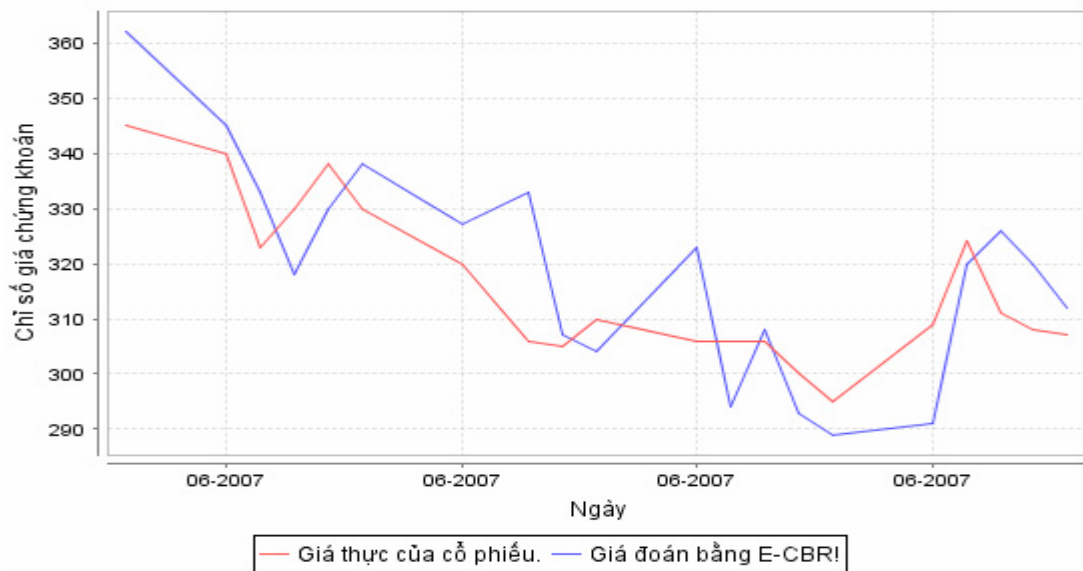
4.2.3.2. Đề xuất phương pháp cải tiến trong tương lai

Ban đầu khi đề xuất bài toán, với suy nghĩ phương pháp thực sự hợp lý, nhưng cuối cùng thực sự kết quả lại không tốt như mong muốn ban đầu. Tuy nhiên, vẫn giữ cùng nhận định ban đầu. Những chỉ số kỹ thuật chính là tín hiệu cho các nhận định nên mua hay nên bán. Vì vậy có lẽ thay vì dự đoán giá cả thì thay vào đó thực hiện dự đoán tín hiệu mua bán xảy ra cũng là một ý tưởng cho việc xây dựng những kết quả tiếp theo.

4.2.4. Phương pháp ECBR

4.2.4.1. Một số kết quả đạt được

Mô hình ECBR như đã đề cập, nó tương tự việc kết hợp nhiều CBR khác nhau nhằm đưa ra một mô hình tối ưu nhất cho bài toán. Một phần vì mô hình này sử dụng khá nhiều công thức tính toán nên tốc độ có hơi chậm. Do đó không thể thực hiện kiểm định những khoảng thời gian dài như các phương pháp ở trên. Ở đây ta thực hiện đánh giá những khoảng thời gian ngắn hơn trên vài cổ phiếu đại diện là ABT và FPT.



Hình 35 : Dự đoán cổ phiếu FPT bằng E-CBR (01/06/2007 - 01/07/2007)

Bảng kết quả trên vài cổ phiếu sử dụng E-CBR:

Tên cổ phiếu	Khoảng thời gian	RMSE	MAPE	MAE
ABT	01/06/2007-01/07/2007	8.34	3.36	4.675
FPT	01/06/2007-01/07/2007	9.7	3.17	7.0
BMC	01/06/2007-01/07/2007	15.9	2.58	13.45

Vì thuật toán đòi hỏi một thời gian tính toán rất lâu khi thực hiện kiểm tra trên một khoảng thời gian nên ở đây chỉ đưa ra một số những kết quả nhỏ. Đối với những cổ phiếu có khối lượng dữ liệu quá khứ lớn thì thuật toán này so với mạng nơron là không thể so sánh được.

4.2.4.2. Đánh giá phương pháp

Phương pháp này đạt không đạt được hiệu quả tốt hơn so với thuật toán mạng nơron, tuy kết quả không thực sự tốt như kì vọng xong nó cũng là một bước trong quá trình

ứng dụng một phương pháp khá mới (sử dụng kết hợp nhiều CBR cho xây dựng mô hình tối ưu) vào dự đoán giá cổ phiếu.

Ưu điểm

- Không cần quá nhiều dữ liệu cũng như thời gian học cho mô hình.
- Trường hợp nếu có một tình huống đã xảy ra trong quá khứ, tương tự như tình huống hiện tại thì mô hình sẽ cho kết quả rất tốt.

Khuyết điểm

- Thuật toán đòi hỏi một thời gian khá lâu trong giải quyết bài toán, đây chính là khuyết điểm lớn nhất.
- Thuật toán có những trường hợp quá khớp xảy ra sẽ đưa ra những kết quả không thực tốt.
- Việc xác định các cửa sổ học và các cửa sổ đánh giá không phải là đơn giản, và đòi hỏi phải có những tìm hiểu sâu hơn.

4.2.4.3. Đề xuất thuật toán cải tiến trong tương lai

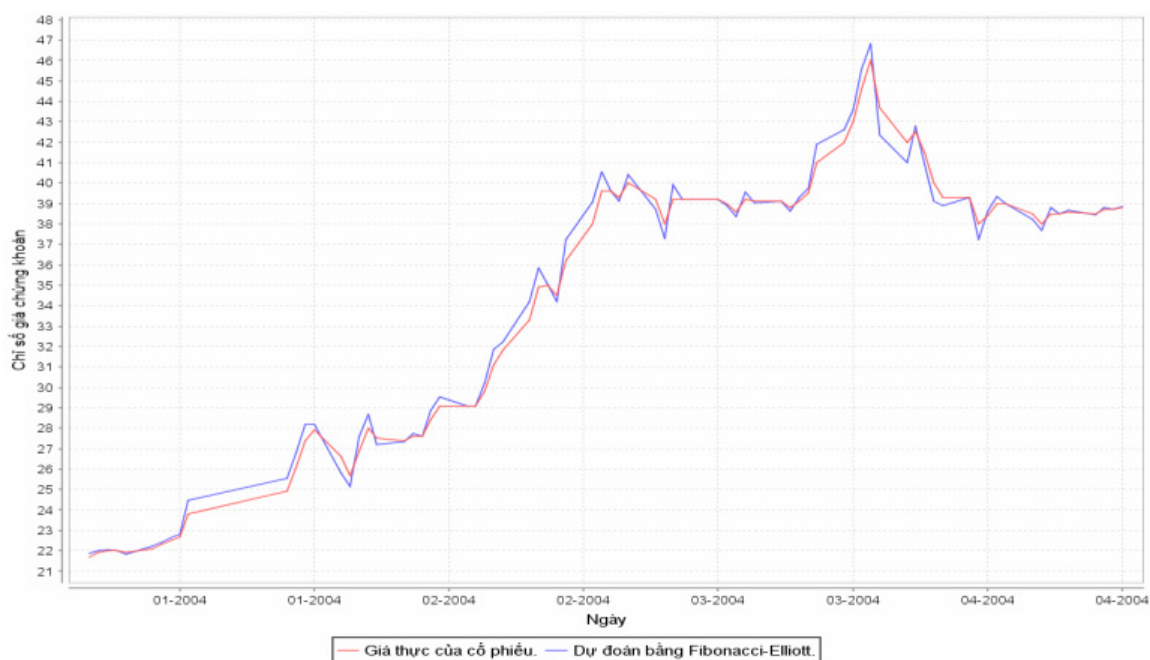
Phương pháp này đòi hỏi một khối lượng các phép tính khá lớn vấn đề thời gian và một vấn đề nan giải. Việc giải quyết, tối ưu về thời gian cho thuật toán có lẽ là một định hướng tiếp theo. Cần có những cải cách trong quá trình xác định các chỉ số cũng như quá trình đối sánh giữa các tình huống sao cho thật hiệu quả. Hiện giờ vẫn chưa có những ý tưởng mới cho việc cải tiến tốc độ thuật toán cũng như hiệu quả của nó. Có lẽ cần một khoảng thời gian tiếp theo để có thể đạt được định hướng này.

4.2.5. Phương pháp Fibonacci-Elliott

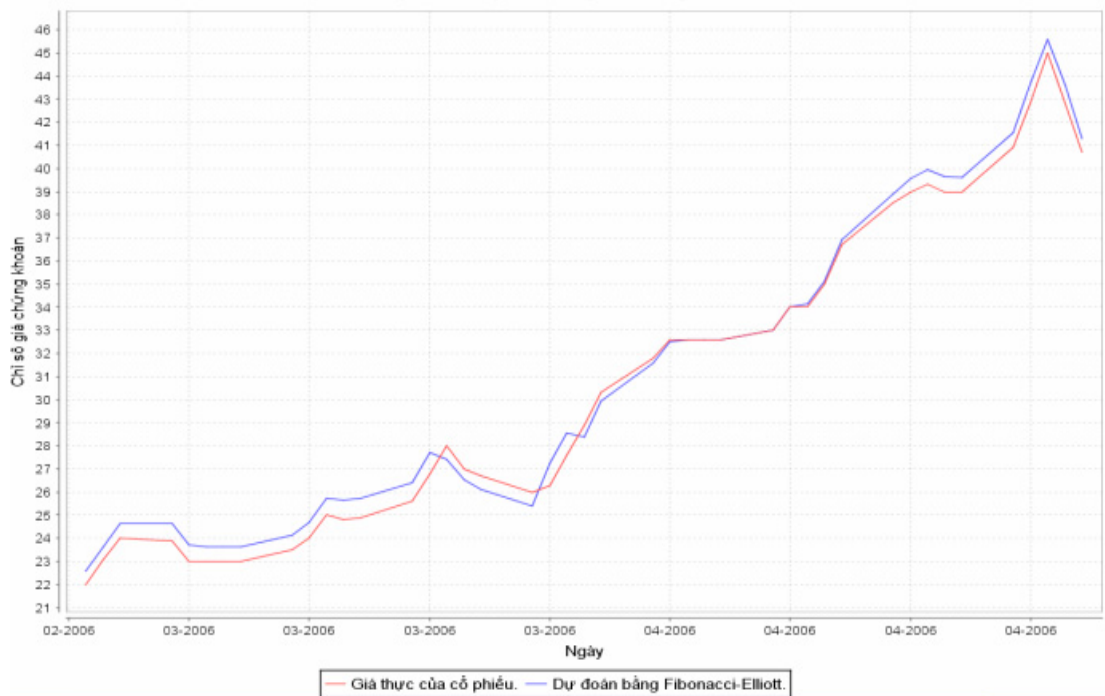
4.2.5.1. Một số kết quả đạt được

Đây có lẽ là một chút kết quả khá tốt trong quá trình xây dựng khóa luận, nó cũng chính là thuật toán dự đoán đạt được kết quả tốt nhất, với hệ số lỗi và đạt khả năng dự đoán chính xác nhất. Một cải tiến nhỏ đã đạt được một số kết quả. Tuy nhiên, thuật

toán chỉ chạy tốt trên một khoảng thời gian ngắn hơn 3 tháng, trong khoảng thời gian này thì nó tốt hơn nhiều so với mạng nơron. Nhưng với khoảng thời gian dài hơn thì thuật toán chạy không hiệu quả. Một điều khá hay là khi thực hiện thuật toán thì với phương pháp trên thì nó cũng cho chúng ta đến một nhận xét rằng, những giá đóng cửa của hầu hết các cổ phiếu trong một khoảng thời gian ngắn nào đó bất kì (hoặc một chu kì ngắn) nào đó luôn có một mối quan hệ tương quan. Như trong thuật toán đó chính là mối quan hệ tương quan hồi qui tuyến tính, mà trước đây không biết có ai đó đã nhận ra được điều này hay chưa? Qua thuật toán cũng như những nhận định ban đầu về thị trường chứng khoán Việt Nam, với những kết quả đạt được theo phương pháp thì bản thân cho rằng, giữa giá cổ phiếu của những ngày trước đó và giá cổ phiếu cần dự đoán luôn có một mối quan hệ nào đó ẩn bên trong và biến đổi theo qui luật trong một khoảng thời gian nhất định nào đó. Và để xây dựng những thuật toán hiệu quả cần có những phương pháp xác định một cách rõ ràng những mối quan hệ này. Và thực sự hầu hết những phương pháp hiện nay đều cố gắng làm như vậy. Một số kết quả được trình bày dưới đây:



Hình 36 : Dự đoán cổ phiếu SAM bằng Fibonacci (1/2004 – 4/2004)



Hình 37 : Dự đoán cổ phiếu CAN bằng Fibonacci (2/2006 – 4/2006)

Một số thống kê nhỏ khi thực hiện dự đoán trên một số cổ phiếu:

Tên cổ phiếu	Khoảng thời gian	RMSE	MAPE	MAE
ABT	13/07/2007-19/07/2007	1.52	1.124	1.193
BMC	13/07/2007-19/07/2007	4.298	0.878	3.674
BTC	13/07/2007-19/07/2007	1.641	4.807	1.443
DMC	13/07/2007-19/07/2007	6.70	3.292	4.467
FPT	13/07/2007-19/07/2007	4.284	1.287	3.763
GMC	13/07/2007-19/07/2007	1.617	2.888	1.555
GMD	13/07/2007-19/07/2007	1.103	0.606	0.857

IMP	13/07/2007-19/07/2007	1.224	0.6794	0.990
LBM	13/07/2007-19/07/2007	1.716	2.159	1.308
PGC	13/07/2007-19/07/2007	0.376	0.5279	0.295
REE	13/07/2007-19/07/2007	1.926	0.989	1.469
RHC	13/07/2007-19/07/2007	0.273	0.323	0.147
SAM	13/07/2007-19/07/2007	1.103	0.5957	0.924
SGC	13/07/2007-19/07/2007	0.237	0.384	0.171
SSC	13/07/2007-19/07/2007	0.661	0.818	0.574
STB	13/07/2007-19/07/2007	0.850	1.126	0.700
TNA	13/07/2007-19/07/2007	1.493	2.021	1.077
VF1	13/07/2007-19/07/2007	0.165	0.457	0.1387
VNM	13/07/2007-19/07/2007	0.979	0.5206	0.929
VSH	13/07/2007-19/07/2007	0.396	0.603	0.3692
VTA	13/07/2007-19/07/2007	0.753	2.532	0.649

Bảng 15 : Bảng tính hệ số lỗi của phương pháp Fibonacci

So sánh chỉ số lỗi RMSE và MAPE với các phương pháp khác ta thấy được sự chênh lệch hệ số lỗi trong dự đoán như thế nào. So với các phương pháp khác, thì phương pháp này thực sự đạt mức độ tốt hơn rất nhiều.

Trong hai nghiên cứu tương tự [29],[47] thì hệ số lỗi đạt được trong bài toán này tốt hơn nhiều. Nhưng đây cũng chỉ là những so sánh mang tính tương đối vì khóa luận sử dụng giá của thị trường chứng khoán trong khi những bài báo này sử dụng một cơ sở dữ liệu khác.

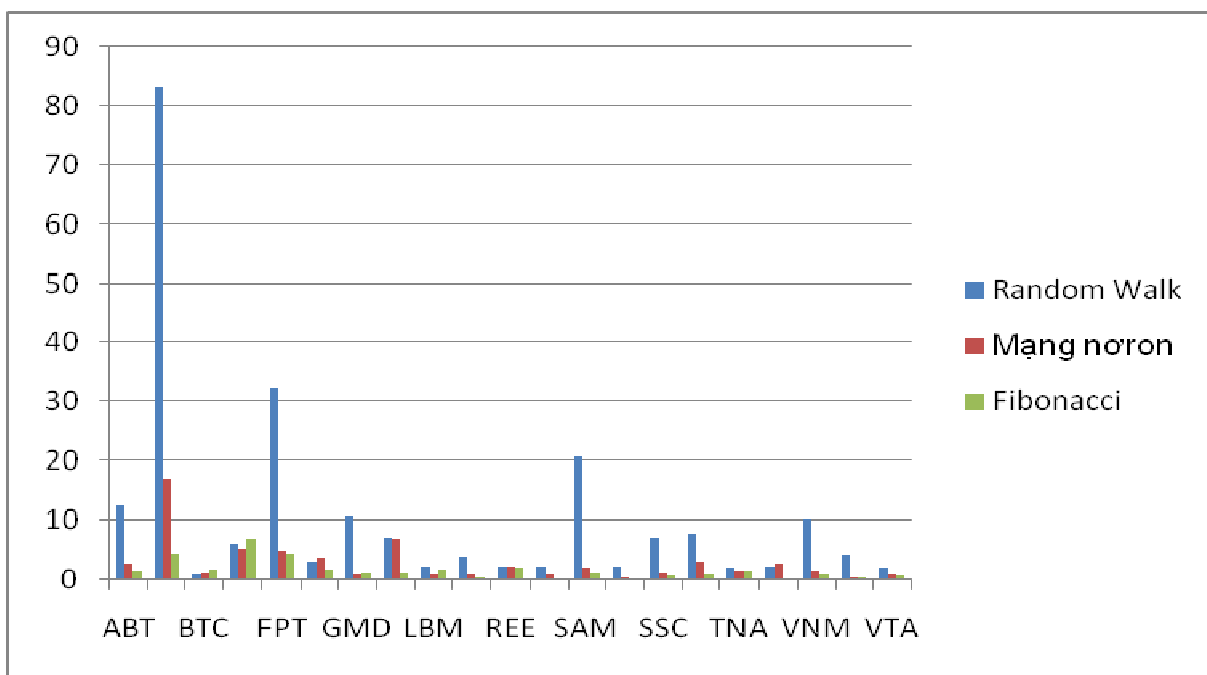
4.2.5.2. Đánh giá thuật toán

Một trong những vấn đề được đề cập và là nền tảng của bài toán là việc áp dụng Fibonacci vào nhằm ước lượng các tham số sao cho đạt được hiệu quả tốt nhất. Có khá nhiều phương pháp ước lượng. Ý tưởng ban đầu là sử dụng mạng nơron 3 lớp. Xong, kết quả không được ưng ý, vì vậy đã đơn giản hóa việc ước lượng bằng phương pháp đơn thuần cho 2 vòng lặp chạy hai biến alpha và beta trong khoảng $[-2, 2]$, mỗi bước nhảy là 0.1. Các giá trị này tùy vào mỗi ngữ cảnh sẽ có các giá trị khác nhau. Xác định được 2 giá trị này chính càng chính xác thì việc dự đoán sẽ cao hơn.

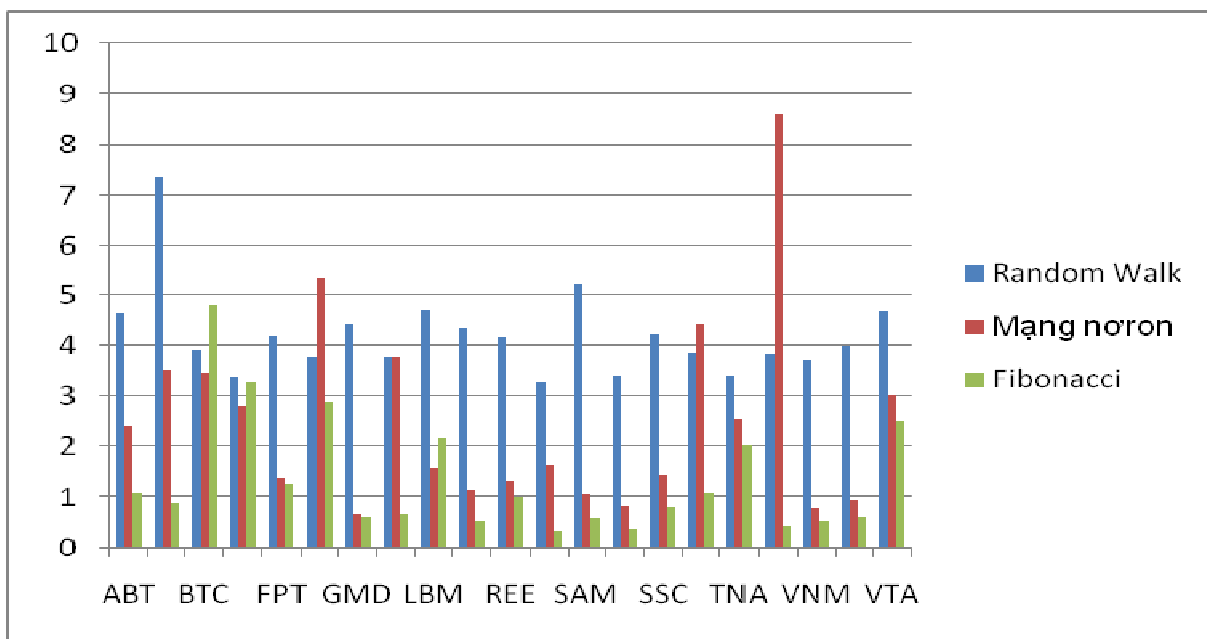
Thuật toán cũng đòi hỏi một khoảng thời gian đủ lớn để có thể thiết lập được bảng quan hệ giữa các biến mờ, với những cổ phiếu có thời gian niêm yết ngắn thì thuật toán sẽ không đạt được hiệu quả cao như mong đợi.

4.2.6. So sánh kết quả đạt được của một số thuật toán trong phân tích kĩ thuật.

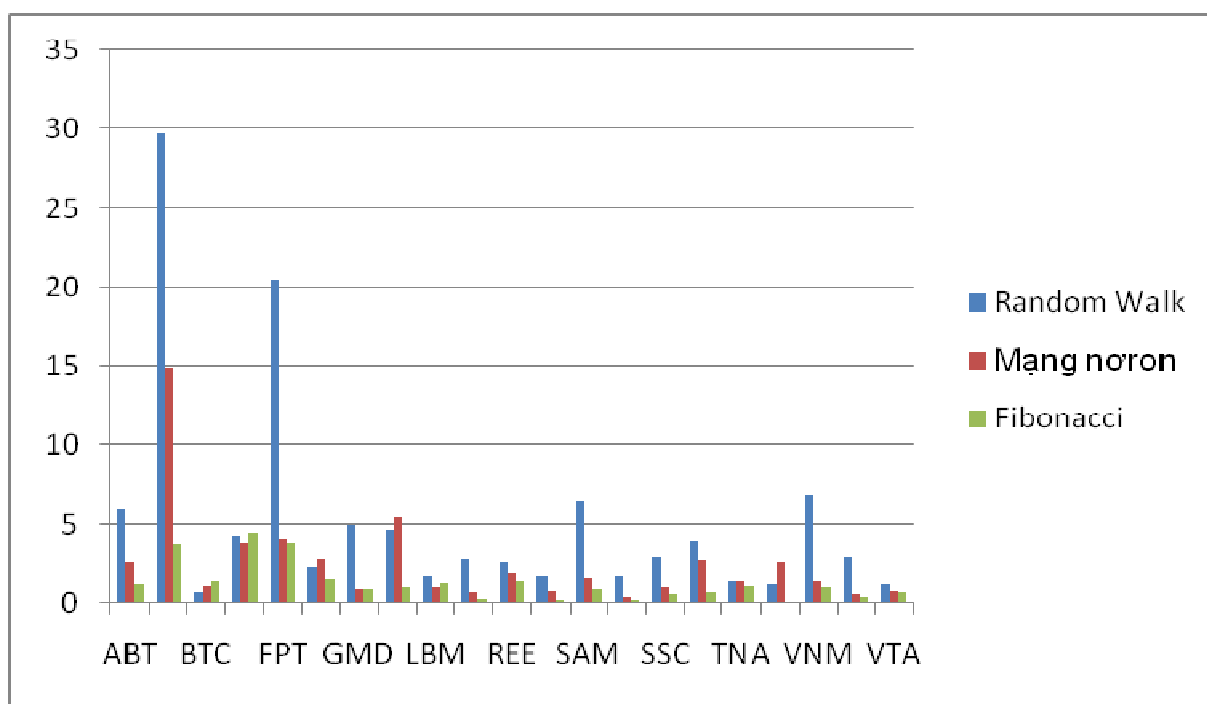
Dựa vào bảng kết quả trên việc đánh giá một số cổ phiếu, ta thấy thuật toán mạng nơron và fibonacci là hai phương pháp khá tương đương nhau. Trên một số những cổ phiếu có khoảng thời gian niêm yết khá lâu mạng nơron chạy rất hiệu quả. Ngược lại phương pháp fibonacci chạy tốt hơn trong các cổ phiếu có thời gian niêm yết ngắn. Một trong hai khác biệt nữa dễ dàng được nhận ra là thuật toán mạng nơron ít ảnh hưởng bởi dữ liệu nhiễu hơn và không bị ảnh hưởng nhiều về hệ số lỗi khi có một sự biến đổi đột ngột rất lớn trong giá của cổ phiếu, trong khi đó thuật toán Fibonacci lại bị ảnh hưởng khá nhiều bởi vấn đề này. Thường thì trong dự đoán giá phương pháp Fibonacci có phần chậm hơn so với giá thực trong một khoảng thời gian nhỏ nào đó. Ngoài ra với thời gian dự đoán ngắn thì phương pháp Fibonacci lại đạt được ưu thế hơn nhiều so với phương pháp mạng nơron. Dưới đây là bảng đánh giá kết quả của hai thuật toán trên, ngoài ra ta thêm phương pháp RW để so sánh mức độ hiệu quả hơn của hai phương pháp này.



Hình 38 : Biểu đồ so sánh lỗi RMSE của 3 phương pháp



Hình 39 : Biểu đồ so sánh lỗi MAPE của 3 phương pháp



Hình 40 : Biểu đồ so sánh lỗi MAE của 3 phương pháp

Trong phần này xin trình bày một số kết quả trong quá trình dự đoán những giá trị thực của thị trường. Một khoảng thời gian quát sát những giá trị trên thị trường với kết quả khá khả quan. Những kết quả đạt được sẽ được thể hiện qua một vài cổ phiếu đại diện là REE và FPT. Khoảng thời gian quan sát (01/07/2007 – 12/07/2007). Đây chỉ là một ví dụ đơn giản trong quá trình dự đoán chuỗi giá thực trên thị trường.

Ngày dự đoán	Mạng nơ-ron	Fibonacci	Giá trị thực
02/07/2007	160	149	144
03/07/2007	160	144	141
04/07/2007	141	141	148
05/07/2007	145	147	143
06/07/2007	155	143	145

09/07/2007	142	145	144
10/07/2007	147	144	151
11/07/2007	153	148	152
12/07/2007	146	150	149

Hình 41 : Bảng kết quả trên cổ phiếu REE

Ngày dự đoán	Mạng nơron	Fibonacci	Giá trị thực
02/07/2007	173	157	157
03/07/2007	173	159	151
04/07/2007	147	159	158
05/07/2007	151	158	158
06/07/2007	158	158	157
09/07/2007	147	158	156
10/07/2007	165	156	159
11/07/2007	153	159	159
12/07/2007	154	159	157

Hình 42 : Bảng kết quả trên cổ phiếu SAM

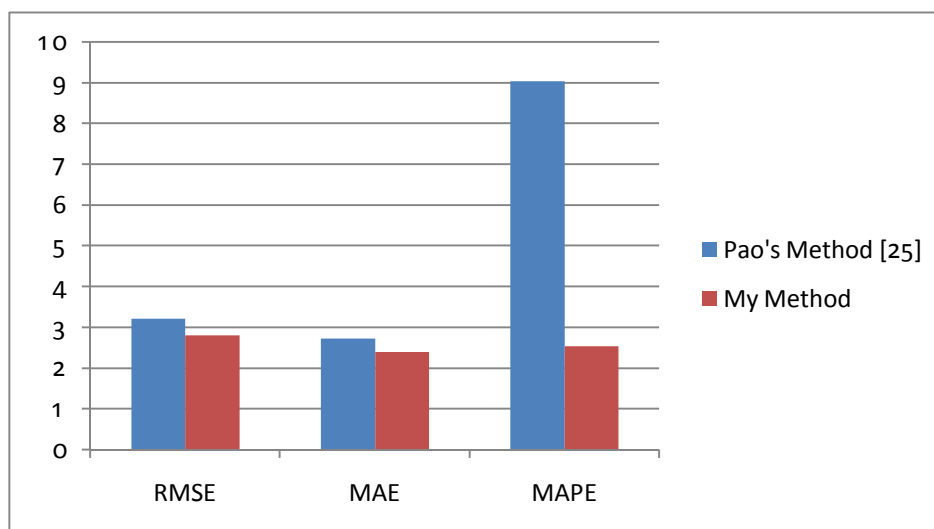
Ngày dự đoán	Mạng nơron	Fibonacci	Giá trị thực
02/07/2007	44	44	43.2
03/07/2007	44	43	42
04/07/2007	43	42	44.1
05/07/2007	44	45	44
06/07/2007	45	44	44.5
09/07/2007	45	45	45
10/07/2007	45	45	45

11/07/2007	46	45	44
12/07/2007	45	44	45

Hình 43 : Bảng kết quả trên cổ phiếu SGC

4.2.7. So sánh với một số công trình khác

Phương pháp mạng nơon so với công trình của Pao [25]



Hình 44 : So sánh với công trình của Pao

Phương pháp Fuzzy time-series so với một số những công trình khác

Trong phương pháp so sánh này thì sử dụng 2 cổ phiếu có thời gian niêm yết đầu tiên nhất trên thị trường chứng khoán thành phố Hồ Chí Minh. Dưới đây là bảng kết quả đạt được:

Mô hình	Thời gian	RMSE
Chen's model [40]	04/01/2003 – 31/12/2004	12.77
Yu's model [26]	04/01/2003 – 31/12/2004	6.72
Proposed model	01/01/2006 – 01/07/2007	REE 2.89
		SAM 2.52

Chương 5. KẾT LUẬN

Như những định hướng ban đầu đã đề ra kết quả đạt được cũng đã giải quyết được một phần. Hai hướng phân tích đều được áp dụng vào giải quyết vấn đề bài toán và đạt được một số hiệu quả nhất định.

Hướng dự đoán thứ nhất được giải quyết bằng kỹ thuật CBR kết hợp Logic mờ và mạng nơron. Bài toán đã được xây dựng trên nền tảng những thuộc tính, những chỉ số mà được sử dụng khá nhiều và hết sức cơ bản để đánh giá mức độ tốt của một cổ phiếu mà hiện nay thường được sử dụng trên thị trường. Những kết quả đạt được cũng chỉ ở mức độ chấp nhận được. Bởi thực sự bộ cơ sở dữ liệu áp dụng cho hệ CBR cũng chưa thực sự do một chuyên gia nào có kinh nghiệm nhiều trong lĩnh vực chứng khoán đề xuất. Mà dữ liệu được xây dựng dựa trên việc thu thập những ý tưởng cũng như những ý kiến của một số nhà đầu tư tham gia thị trường chứng khoán được một thời gian. Để hệ thống có thể thể thực thi hiệu quả hơn thì vấn đề này là cực kì quan trọng và đòi hỏi phải đầu tư khá nhiều công sức. Nhưng vì điều kiện chưa cho phép nên cũng thực sự là một điều khá tiếc nuối. Hi vọng trong thời gian sắp tới sẽ có những bước phát triển trong bộ cơ sở dữ liệu chứng khoán trong phân tích đánh giá cổ phiếu này.

Một hướng tìm hiểu khác là sử dụng phân tích kỹ thuật vào giải quyết bài toán dự đoán giá của một cổ phiếu. Phần này được tìm hiểu khá nhiều phương pháp khác nhau, mỗi phương pháp đạt được một số những kết quả nhất định còn một số cũng chỉ ở những bước đầu thử nghiệm.

Phương pháp mạng nơron: Áp dụng 2 kiểu áp dụng khác nhau nhưng mức độ bài toán chưa thực sự đạt được hiệu quả thật cao. Trong một số đề xuất được đưa ra trong những phương pháp này cũng chỉ là những ý tưởng ban đầu. Kết quả cần nhiều thời gian hơn trong nhận định kết quả đạt được, cũng như kiểm chứng sự phù hợp của mô hình áp dụng vào thị trường chứng khoán Việt Nam.

Phương pháp ECBR: Kết quả đạt được cũng chỉ ở mức độ thử nghiệm. Một nhược điểm khá lớn cho thuật toán này là đòi hỏi những tính toán khá nhiều. Khi thực thi ứng dụng thì cần một khoảng thời gian khá lâu, có thể đây là một nhược điểm khá lớn cần được cải tiến sau này. Tuy nhiên, thuật toán này không cần một lượng dữ liệu quá nhiều. Do đó nó có thể áp dụng trên những chứng khoán mà chỉ lên sàn giao dịch trong một khoảng thời gian ngắn. Đây là bài toán mà mạng nơron không giải quyết được.

Phương pháp sử dụng time-series model kết hợp Fibonacci: Trong tất cả các phương pháp dự đoán kỹ thuật thì thuật toán này đạt được hiệu quả cao nhất trong quá trình dự đoán. Dựa trên mô hình Framework chuẩn đã được xây dựng, với một số những cải tiến nhỏ nhưng hiệu quả đạt được là khá bất ngờ ngoài dự kiến. Tuy nhiên, để thuật toán có thể đi vào ứng dụng thực tế thì còn phải cần một thời gian khá dài để kiểm tra tính đúng đắn và phù hợp với thực tế thị trường chứng khoán Việt Nam, vì thực tế thì kết quả đạt được cũng chưa phải là quá cao.

Khóa luận và ứng dụng sẽ được tiếp tục phát triển sau này, bằng cách đầu tư nhiều thời gian hơn trong những lý thuyết chứng khoán để có những kiến thức sâu hơn, đồng thời cũng cũng có một số những biến đổi hợp lý hơn cho ứng dụng có thể phù hợp với ngữ cảnh của thị trường chứng khoán Việt Nam. Việc dự đoán giá đóng cửa vào ngày hôm sau thì chỉ là những bước đầu trong bài toán dự đoán chứng khoán, một trong những bước phát triển trong tương lai có thể áp dụng vào bài toán dự đoán chứng khoán có thể dự đoán khoảng thời gian dài hơn, việc dự đoán khoảng thời gian dài hơn có thể sẽ có lợi hơn nhiều trong vấn đề giúp nhà đầu tư dự đoán chứng khoán. Vì vậy thay vì những bước phát triển trong dự đoán giá hôm sau thì sẽ có thể phát triển lên một thuật toán mới hỗ trợ dự đoán thời gian dài hơn 5 – 10 ngày chẳng hạn.

TÀI LIỆU THAM KHẢO

Tài liệu trong nước

- [1] Đào Công Bình, Minh Đức. *Hướng dẫn đầu tư cổ phiếu*. Nhà xuất bản trẻ. 2007
- [2] Dữ liệu chứng khoán trực tuyến www.hsc.com.vn
- [3] <http://phantichcophieu.com>
- [4] Lê Hoàng Thái, SVTH: Lê Trọng Ngọc. *Hệ CBR cho chuẩn đoán bệnh tim HCM* : s.n., 2003
- [5] Nguyễn Anh Dũng, Tạ Văn Hùng. *Thị trường chứng khoán (Dành cho người mới bắt đầu)*. HCM : Nhà xuất bản thống kê, 2001
- [6] Nguyễn Thanh Tuyền, Nguyễn Đăng Dờn. *Thị trường chứng khoán (Securities Market)*. Nhà xuất bản thống kê 2005.
- [7] Nguyễn Thanh Tuyền, Nguyễn Đăng Dờn. *Thị trường chứng khoán*. HCM : Nhà xuất bản thống kê, 2005
- [8] Quang Ngọc. *Làm giàu từ thị trường chứng khoán Việt Nam*. Nhà xuất bản văn hóa thông tin. 2007
- [9] Trần Hạnh Nhi, Trần Minh Triết. SVTH: Nguyễn Thanh Bình, Nguyễn Thiện Nhân. *Xây dựng môi trường hỗ trợ giao dịch chứng khoán ảo tại Việt Nam*. HCM : s.n., 2003
- [10] Trang web về các chỉ số ngành. <http://www.cbv.vn>
- [11] Trang web về thông tin các chỉ số EPS, P/E, P/B của một số cổ phiếu <http://www.tvsi.com.vn> , www.vietstock.com.vn
- [12] Trang web về thông tin công bố của các công ty www.bsc.com.vn
- [13] Trang web với các kiến thức về tài chính. <http://www.saga.vn/default.aspx>
- [14] Võ Thành Long, Nguyễn Quang Hải. *Chứng khoán để đầu tư thành công*. Nhà xuất bản thanh niên 2007.

Tài liệu ngoài nước

- [15] Achelis, Steven B. *Technical Analysis from A to Z*. s.l. : Equis International, 2003
- [16] Aamodt, Agnar. *Case-Based Reasoning: Foundation Issue, Methodological Variations, and System Approaches*. Catalonia : AI Communication, 1994
- [17] Berk, Richard. *An Introduction to Ensemble Methods*. Los Angeles : UCLA.
- [18] Buckley, William siler & Jame J. *Fuzzy Expert Systems and Fuzzy Reasoning*. New Jersey : Wiley, 2005
- [19] C.K.Shiu, Sanka K.Pal & Simon. *Foundations of Soft Case-Based Reasoning*. New Jersey : Wiley, 2004
- [20] Các trang web về phân tích kỹ thuật. <http://www.fmlabs.com> , <http://stockcharts.com> , <http://www.fxwords.com>
- [21] Dorado, Juan R. Rabunal and Julian. *Artificial Neural Network in real life applications*. Hershey : Idea Group, 2006
- [22] Friedman-Hill, Ernest *The Rule Engine for the Java™ Platform*. s.l. : Sandia National Laboratories
- [23] Gupta, Kate Smith & Jatinder. *Neural Networks in bussiness: Techniques and Applications*. Chocolate Avenue Hershey : Idea Group, 2002
- [24] Haecke, Bernard Van. *JDBC 3.0 Java Database Connectivity*. 909 Third Avenue New York : M&T, 2002
- [25] Hsiao-Tien Pao. *Forecasting electricity market pricing using artificial neural network*. Taiwan : Elsevier, 2007
- [26] Hui-Kuang Yu, *Weighted fuzzy time series models_ for _TAIEX forecasting* . Elsevier 2006.
- [27] Januskevvcus, Marius. *Testing stock market Efficiency Using Neural Networks* Riga : SSE Riga Working paper, 2003

- [28] JingTao YAO, Chew Lim TAN. *Guidelines for Financial Forecasting with Neural Networks*
- [29] Kunhuang Huarng, Hui-Kuang Yu. Taiwan : *A Type 2 fuzzy time series model for stock index forecasting*. Elsevier, 2005
- [30] Leung, James N.K. Liu & Tommy T.S. *A Web-based CBR Agent for Financial Forecasting*. Hong Kong : Department of Computing, Hong Kong Polytechnic University, 2002
- [31] *Linear Recurrence Equation*. <http://mathworld.wolfram.com/>
- [32] Marijana Zekic, MS. *Neural Network Applications in Stock Market Predictions - A Methodology Analysis*. Osijek : Faculty of Economics Osijek
- [33] McMillan, david G. *Non-linear forecast of stock returns: Does volume help?* St. Andrew : Elsevier, 2007
- [34] Onstine, Richard Hightower & Warner. *Professional Java Tools for Extreme Programming*. Indianapolis : Wiley, 2004
- [35] Ramyar, Roy Batchelor and Richard. *Magic numbers in the Dow*. London : Cass Business School, 2006.
- [36] Raquel Gareta, Luis M. Romeo , Antonia Gil. *Forecasting of electricity prices with neural networks*. Zaragoza : Elsevier, 2005
- [37] Riordan, Bjarne K.Hansen & Denis. *Weather Prediction Using Case-Based Reasoning and Fuzzy Set Theory*. Nova Scotia, Canada : Dalhousie University, 2001
- [38] Santanu Chaudhury, Tuhina Signh & Partha S. Goswami. *Distributed fuzzy case based reasoning* . New Delhi : Elsevier, 2004
- [39] Se-Hak Chun, Yoon-Joo Park. *A new hybrid data minng technique using regression case base reasoning: Application to financial forecasting*. Seoul. Elsevier 2005.

- [40] Shyi-Ming Chen, Forecasting enrollments based on fuzzy time series. Elsevier 1995.
- [41] *The JFreeChart Class Library Version 1.0.5*. Source Open. 2000-2007
- [42] *Thư viện phân tích kỹ thuật TALib*. License, BSD. s.l. : <http://ta-lib.org/>
- [43] W.M.Kwong, James N.K. Liu & Raymon. *Automatic extraction and indentification of chart pattern towards financials forecast*. Hong Kong : Elsevier, 2006
- [44] Watson, I. *Case-based reasoning is a methodology not a technology*. Salford : Elsevier, 1999
- [45] Watson, Ian. *Apply case-based reasoning: Techniques for Enterprise Systems*. San Francisco, California : Morgan Kaufmann, 1997
- [46] Watson, Ian. *Case-Based Reasoning Application for Engineerings*.l. : University of Auckland
- [47] Yu, Hui-Kuang. A refined fuzzy time-series model for forecasting. Taiwan : Elsevier, 2004

PHỤ LỤC

➤ Mô hình ARIMA

Giới thiệu: ARIMA là các lớp các mô hình cho dự đoán một chuỗi thời gian (time series) có thể được thiết lập “tĩnh hóa” qua sự biến đổi sai khác (differencing) hoặc biến đổi logarithm nhằm thiết lập mối quan hệ giữa các giá trị trong chuỗi thời gian để giảm được hệ số lỗi trong quá trình dự đoán.

ARIMA là viết tắt của “Auto-Regressive Integrated Moving Average”. Trong phương trình dự đoán “Auto-Regressive” thường là thể hiện sự khác biệt hay độ trễ của một chuỗi được thực hiện biến đổi sai khác “differencing”. Độ trễ trong lỗi dự đoán thường được gọi là trung bình chuyển động (Moving Average). Và các chuỗi thời gian cũng cần được thực hiện thực hiện phép biến đổi sai khác (differencing) nhằm tạo sự “ổn định” hay “sự không thay đổi” được gọi là một phiên bản được tổng hợp của chuỗi ổn định (stationary series). Ví dụ, Đường đi ngẫu nhiên (Random-walk) và các mô hình hướng ngẫu nhiên (random-trend), các mô hình “autoregressive” và các mô hình làm trơn lũy thừa (exponential smoothing) (vd. Exponential weighted moving averages) là tất cả các trường hợp của mô hình ARIMA.

Mô hình ARIMA thường được gọi là $ARIMA(p,d,q)$ với các giá trị p,d,q như sau:

- p là một số “autoregressive”.
- d là một số “nonseasonal difference”.
- q là một số thể hiện độ chậm của lỗi dự đoán trong đẳng thức dự đoán.

Để xác định một ARIMA thích hợp cho chuỗi thời gian, chúng ta cần xác định tuần tự “sự sai khác” (differencing) cần được “ổn định hóa” (stationarize) các chuỗi và bỏ đi các thuộc tính không ổn định hoặc dư thừa. Có lẽ, mối tương quan giữa những biến đổi - ổn định (variance-stabilizing) như: logging hoặc deflating. Nếu chúng dừng ở đây và dự đoán thì chuỗi sau khi được làm khác (differenced) là hằng số.

Việc quan trọng là xác định một mô hình ARIMA thích hợp và hợp lý cho bài toán mà chúng ta đang tiếp cận.

ARIMA(0,1,0) = random-walk: Trong những mô hình mà chúng ta đã tìm hiểu ở trước, chúng ta gặp phải 2 chiến lược cho việc rút ra “autocorrelation” trong lỗi dự đoán. Một hướng tiếp cận, mà đầu tiên chúng ta sử dụng phân tích hồi quy (regression analysis) với các độ trễ thêm vào sao cho chuỗi sau khi biến đổi của ta được ổn định. Ví dụ, đối với bài toán mô hình phát triển random-walk cho một chuỗi thời gian Y . Đẳng thức dự đoán cho mô hình này có thể được viết như sau:

$$\hat{Y}(t) - Y(t-1) = \mu$$

Ở đây hằng số là μ là độ sai khác trung bình trong Y . Điều này có thể được xem như một mô hình hồi quy được tạo, trong đó $\text{DIFF}(Y)$ là một biến phụ thuộc và không có một biến độc lập nào khác hơn hằng số này. Bởi vì, nó bao gồm duy nhất một “nonseason difference” và một hằng số. Nó được phân loại như một mô hình $\text{ARIMA}(0,1,0)$ với hằng số.

ARIMA(1,1,0) = difference first-order autoregressive model: Nếu các lỗi của mô hình random-walk là autocorrelated, vấn đề này dễ dàng được giải quyết bằng cách thêm một biến phụ thuộc trong đẳng thức dự đoán. Bằng cách hồi quy $\text{DIFF}(Y)$. Điều này sẽ tạo ra đẳng thức sau:

$$\hat{Y}(t) - Y(t-1) = \mu + \varphi(Y(t-1) - Y(t-2))$$

Hoặc chúng ta có thể sắp xếp lại theo đẳng thức dưới đây:

$$\hat{Y}(t) = \mu + Y(t-1) + \varphi(Y(t-1) - Y(t-2))$$

ARIMA(0,1,1) without constant = simple exponential smoothing: Một chiến lược khác cho việc hiệu chỉnh các hệ số lỗi autocorrelated trong mô hình random-walk được đề xuất là mô hình “simple exponential smoothing”. Trong một số nghiên cứu về cho thấy việc sử dụng một số những quan sát gần đây nhất để thực hiện dự đoán giá trị tương lai không tốt bằng việc sử dụng một số những chỉ số trung bình của một vài quan

sát trong quá khứ để có thể lọc bỏ được nhiễu và ước lượng một cách chính xác hơn giá trị kì vọng trung bình (mean). Mô hình “simple exponential smoothing” sử dụng “exponentially weighted moving average” của các giá trị quá khứ để đạt được hiệu quả này. Công thức dự đoán cho “simple exponential smoothing” có thể được viết bằng một số các phương pháp khác nhau, dưới đây là một ví dụ:

$$\hat{Y}(t) = Y(t-1) - \theta e(t-1)$$

trong đó $e(t-1)$ là lỗi trong khoảng thời gian $(t-1)$ trước đó.

ARIMA(0,1,1) with constant = simple exponential smoothing with growth: Bằng cách vận dụng mô hình SES như một mô hình ARIMA ta thấy thức sự linh động hơn. Trước tiên, hệ số ước lượng $AM(1)$ có thể là số âm: điều này tương ứng như một nhân tố làm trơn (smooth) hơn là 1 trong mô hình SES, điều này thực sự không cho phép trong mô hình SES. Thứ hai, Chúng ta có sự chọn lựa bao gồm giá trị hằng số trong mô hình ARIMA mà ta muốn, để ước lượng xu hướng trung bình khác 0. Mô hình ARIMA(0,1,1) với hằng số có công thức sau:

$$\hat{Y}(t) = \mu + Y(t-1) - \theta e(t-1)$$

Dự đoán giá trị tiếp theo từ mô hình này là định tính tương tự mô hình SES, loại trừ xu hướng hay quỹ đạo trong dự đoán thời gian dài là đường dốc nghiêng hơn là đường nằm ngang.

ARIMA(0,2,1) or (0,2,2) without constant = linear exponential smoothing: Mô hình “linear exponential smoothing” là các mô hình ARIMA mà sử dụng hai sự sai khác trong mối liên kết giữa hai MA. Khác biệt cấp 2 của chuỗi Y là không đơn giản là sự khác biệt giữa Y và chính nó mà bị trễ trong 2 khoảng thời điểm (period). Sự khác biệt cấp 2 thường ở thời điểm t bằng $(Y(t)-Y(t-1)) - (Y(t-1)-Y(t-2)) = Y(t) - 2Y(t-1) + Y(t-2)$.

Mô hình ARIMA(0,2,2) không hằng số dự đoán rằng sự khác biệt cấp 2 của chuỗi bằng hàm tuyến tính của 2 lỗi dự đoán:

$$\hat{Y}(t) - 2Y(t-1) + Y(t-2) = -\theta_1 e(t-1) - \theta_2 e(t-2)$$

Chúng ta có thể viết lại như sau:

$$\hat{Y}(t) = 2Y(t-1) - Y(t-2) - \theta_1 e(t-1) - \theta_2 e(t-2)$$

Ở đây θ_1 , θ_2 là hai hệ số MA(1) và MA(2). Trong mô hình LES thì θ_1 , và θ_2 có công thức như sau:

$$\hat{Y}(t) = 2Y(t-1) - Y(t-2) - 2(1-\alpha)e(t-1) - (1-\alpha)^2 e(t-2)$$

A "mixed" model--ARIMA(1,1,1): Hai tính chất “autoregressive” và “moving average” có thể được trộn lại trong cùng một mô hình. Ví dụ, mô hình ARIMA(1,1,1) với hằng số sẽ có dạng thức dự đoán :

$$\hat{Y}(t) = \mu + Y(t-1) - \phi(Y(t-1) - Y(t-2)) - \theta e(t-1)$$

Tuy nhiên, mô hình kiểu này kết hợp cả hai kiểu trong mô hình thường dẫn đến việc quá khớp dữ liệu và trường hợp hệ số không đồng nhất.

Việc ứng dụng các mô hình ARIMA vào việc chuẩn hóa cũng như dự đoán các giá trị chứng khoán được sử dụng khá nhiều. Ở trên đây chỉ mang tính chất giới thiệu và không quá đào sâu vào kiến thức của các mô hình này. Những kiến thức cơ bản này sẽ được áp dụng vào bài toán.

➤ Một số những chỉ số kĩ thuật

Một số các thông số trong phân tích kĩ thuật được sử dụng trong khóa luận được đề cập ở đây:

Open: Giá mở cửa trong ngày của một cổ phiếu.

High: Giá cao nhất trong ngày.

Low: Giá thấp nhất trong ngày.

Close: Giá đóng cửa trong ngày.

Volume: Số lượng cổ phiếu giao dịch trong.

Accumulation/Distribution: là chỉ số moment liên quan đến sự thay đổi giá và khối lượng giao dịch của cổ phiếu.

A/D thực sự là một biến thể của chỉ số OBV(On Balance Volume). Cả hai chỉ số này đều cố gắng khẳng định những sự thay đổi trong giá bằng cách so sánh khối lượng giao dịch với liên quan đến các chỉ số giá.

Khi chỉ số A/D chuyển động lên, đó là dấu hiệu cổ phiếu đang được tích lũy bởi vì hầu hết khối lượng giao dịch kèm theo đó là giá cả cũng tăng theo. Khi chỉ số này chuyển động xuống. Đó là dấu hiệu cổ phiếu đang bị phân tán. Vì hầu hết không lượng giao dịch liên quan giảm liên quan đến sự chuyển động xuống của giá.

Sự phân kì giữa A/D và giá của cổ phiếu giả sử rằng sắp xảy ra. Khi sự phân kì xảy ra giá thường thay đổi để khẳng định A/D. Ví dụ, nếu chỉ số này đang tăng lên và giá cổ phiếu đang giảm, giá có lẽ sẽ đảo ngược.



Hình 45 : Đồ thị biểu diễn A/D giảm trong khi giá tăng

Công thức A/D:

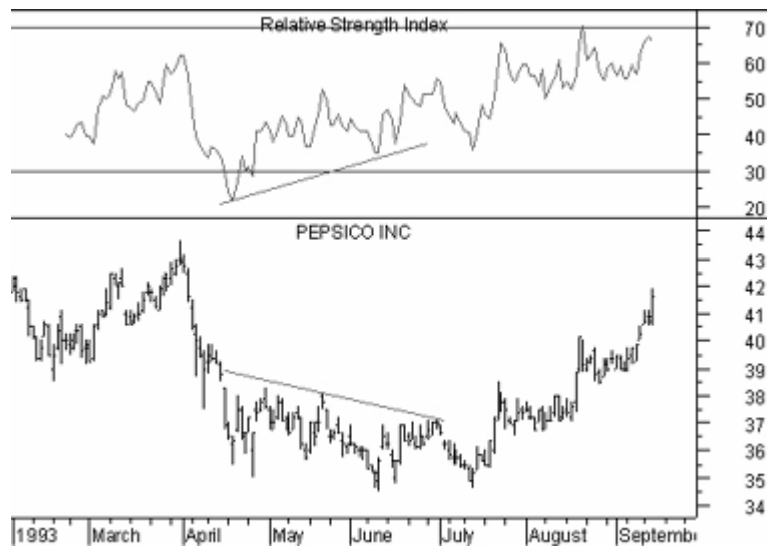
$$\sum \left[\frac{(close - low) - (high - close)}{(high - low)} * volume \right]$$

RSI (Relative Strength Index): Đây là chỉ số khá phổ biến và được sử dụng nhiều trong phân tích kỹ thuật.

Là một sự giao động dựa vào giá có giá trị trong khoảng [0,100]. Phương pháp phổ biến nhất khi phân tích RSI là tìm kiếm sự phân kì mà mỗi cổ phiếu đang tạo ra một giá trị high mới. Nhưng RSI đang giảm xuống dưới so với giá trị cao trước đó của nó. Sự phân kì này là một tín hiệu cho sự đảo ngược giá sắp xảy ra. Khi RSI rớt xuống dưới những nhanh chóng dưới nhiều so với những giá trị gần nó nhất người ta gọi đó là “failure swing”. Đây là tín hiệu chỉ ra sẽ có sự đảo ngược giá sắp xảy ra.

Có 5 cách sử dụng RSI để phân tích biểu đồ giá như sau:

- Cao nhất và thấp nhất (Tops and Bottom): Cao nhất thường là 70 và thấp nhất là 30. Nó thường hình thành những vị trí cao nhất và thấp nhất trước khi biểu đồ giá cơ bản.
- Sự hình thành đồ thị (Chart Formations): RSI thường hình thành các mẫu đồ thị như “head and shoulder” hoặc “triangles” mà không thấy trong đồ thị.
- Gãy cánh (Failure Swings): Được biết như là sự vượt qua mức kháng cự hay sự phá vỡ. Đó chính là lúc mà RSI vượt qua đỉnh trước đó và rớt xuống dưới giá trị thấp gần đó.
- Hỗ trợ và kháng cự (Support and Resistance): RSI chỉ ra một cách rõ ràng hơn so với giá của nó.
- Sự phân kì (Divergences): Như đã trình bày ở trên, sự phân kì xảy ra khi giá tạo ra một đỉnh mới (có thể cao hoặc thấp) mà không khẳng định được bằng một đỉnh mới (cao hoặc thấp) trong RSI. Giá thường đứng và chuyển động theo hướng của RSI.



Hình 46 : Một ví dụ và giá theo hướng của RSI

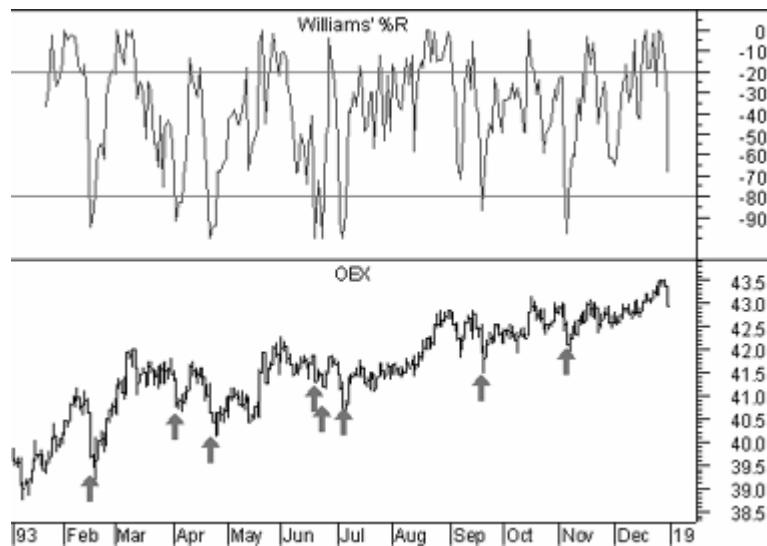
Công thức RSI:

$$RSI = 100 - \left(\frac{100}{1 + \left(\frac{U}{D} \right)} \right)$$

U : Trung bình giá thay đổi tăng.

D : Trung bình giá thay đổi giảm.

William's %R: là chỉ số được sử dụng để đo mức độ vượt mức của hành động mua và bán. %R thường sử dụng giá trị âm (ví dụ, -20%). Với mục đích phân tích và bàn luận nên ở đây ta coi như bỏ qua giá trị âm. Từ 80 – 100% Chỉ ra hành động bán quá mức, trong khi từ 0 – 20% chỉ ra hành động mua quá mức. Dựa trên những tín hiệu ta có thể đợi cho giá thị trường ổn định trở lại đúng hướng trước khi bạn thực hiện giao dịch.

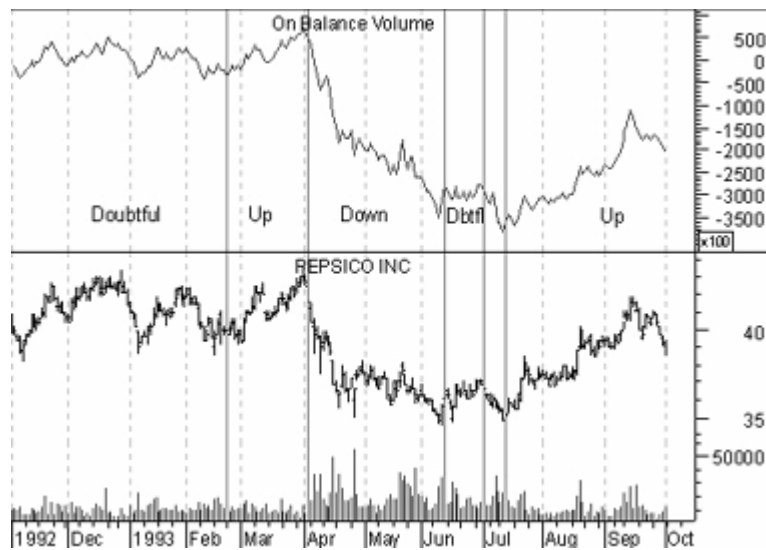


Hình 47 : Biểu đồ thể hiện %R

Công thức %R:

$$\%R = \left\{ \frac{\text{Highest High in } n - \text{period} - \text{Today's close}}{\text{Highest High in } n - \text{period} - \text{Lowest low in } n - \text{period}} \right\} * -100$$

OBV(On Balance Volume): là chỉ số thể hiện sự liên quan giữa khối lượng giao dịch và giá cả. Đây là vấn đề khác phức tạp vì vậy ở đây sẽ không trình bày chi tiết. Mà hiểu đơn giản như sau: “Sự biến đổi tăng giảm của chỉ số OBV là tín hiệu tăng giảm của giá cổ phiếu”.



Công thức được trình bày như sau:

- Nếu giá đóng cửa hôm nay thấp hơn giá đóng cửa hôm trước:

$$OBV = Yesterday's\ OBV + Today's\ Volume$$

- Nếu giá đóng cửa hôm nay cao hơn giá đóng cửa hôm trước.

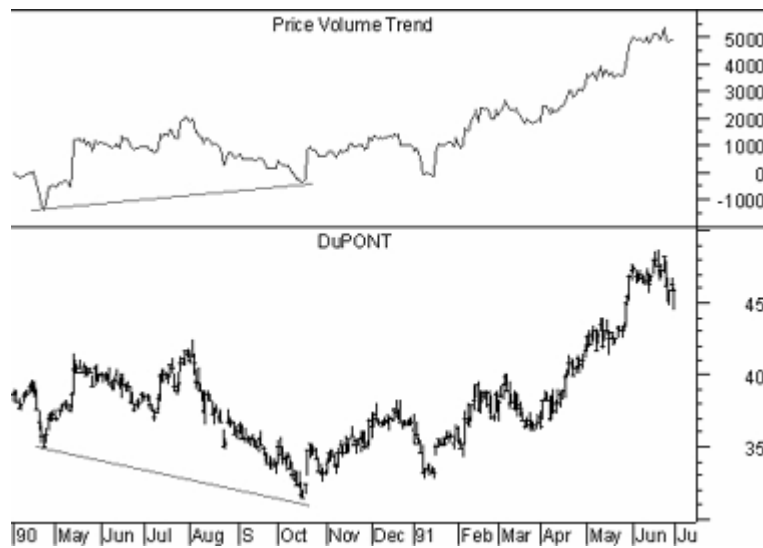
$$OBV = Yesterday's\ OBV - Today's\ Volume$$

- Nếu giá đóng cửa hôm nay bằng giá đóng cửa hôm trước.

$$OBV = Yesterday's\ OBV$$

The Price and Volume Trend: tương tự OBV trong đó nó là tổng của các tích lũy của khối lượng giao dịch mà được điều chỉnh phụ thuộc trên những thay đổi trong giá đóng cửa.

Một số nhà đầu tư cho rằng PVT minh họa luồng tiền trong và ngoài của một cổ phiếu rõ ràng hơn so với OBV.



Hình 48 : Biểu đồ ví dụ về PVT

Công thức tính PVT:

$$\left(\left(\frac{\text{Close} - \text{Yesterday's Close}}{\text{Yesterday's Close}} \right) * \text{Volume} \right) + \text{Yesterday's PVT}$$

Ease of Movement: cung cấp một giá trị thể hiện giá và khối lượng giao dịch cho khoảng thời gian đó. Nó tính toán một cách sự xu hướng mà giá chuyển động. Giá càng tăng thì mức độ giao dịch càng lớn và sự chuyển động dễ dàng hơn.

EMV được tính như sau:

$$EMV = \frac{\text{Midpoint Move}}{\text{BoxRatio}}$$

Trong đó các giá trị Midpoint và BoxRatio được tính như sau:

:

$$\text{Midpoint} = \frac{\text{Today's High} + \text{Today's Low}}{2} - \frac{\text{Yesterday's High} + \text{Yesterday's Low}}{2}$$

$$BoxRatio = \frac{Volume}{High - Low}$$

Stochastic Oscillator (Stock%K và Stock%D): là chỉ số moment mà chỉ ra vị trí của giá trị đóng cửa hiện hành tương quan với khoảng giá trị high/low qua một khoảng thời gian. Các mức độ đóng cửa là gần đỉnh của khoảng chỉ ra độ tích lũy (sức ép mua) và những giá trị gần với khoảng dưới chỉ ra độ phân bổ (sức ép bán).

$$\%K = 100 * \left(\frac{Recent\ Close(n) - Lowest\ Low(n)}{Highest\ Close(n) - Lowest\ Low(n)} \right)$$

$$\%D = MA - 3 - period$$

Trong đó MA-3-period là đường trung bình 3 ngày của %K .

ChangeIn4Close: Đây không phải là một chỉ số kỹ thuật mà là giá trị phần trăm thay đổi của 4 ngày trong tương lai so với ngày hiện hành. Công thức được tính như sau:

$$ChangeIn4Close = \frac{Future4day's\ Close - Today's\ Close}{Today's\ Close}$$

Những kiến thức cơ bản về các chỉ số chỉ trình bày một cách sơ lược, nếu muốn tìm hiểu sâu hơn có thể tìm thấy đầy đủ kiến thức trong [15][20].

➤ Lý thuyết sóng Elliott và chuỗi Fibonacci

Fibonacci trong phân tích kỹ thuật - Dự đoán xu hướng giá chứng khoán

Có một “tỷ lệ” rất đặc biệt được sử dụng để mô tả tính cân đối của vạn vật từ những khối cấu trúc nhỏ nhất của thiên nhiên như nguyên tử cho đến những thực thể có kích thước cực kỳ khổng lồ như thiên thạch. Không chỉ thiên nhiên phụ thuộc vào nó để duy trì sự cân bằng mà thị trường tài chính có vẻ như cũng vận động theo một quy luật tương tự.

Trong phạm vi bài viết, chúng ta sẽ xem qua một vài công cụ phân tích kỹ thuật được phát triển dựa trên các nghiên cứu trên cái mà người ta gọi là “tỷ lệ vàng” này.

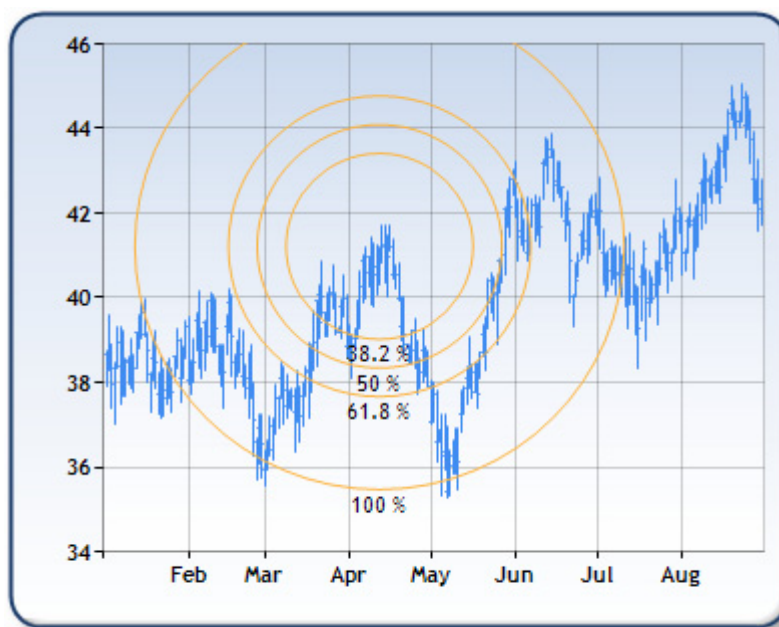
Các nhà toán học, khoa học, và tự nhiên học đã biết đến “tỷ lệ vàng” này trong nhiều năm. Nó được rút ra từ dãy Fibonacci, do nhà toán học người Ý, Leonardo Fibonacci (1175-1250) tìm ra. Trong dãy Fibonacci (1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 610, etc), mỗi số trong dãy là tổng của hai số trước đó. Điều đặc biệt nhất trong dãy này là bất kỳ một số nào cũng đạt giá trị xấp xỉ 1.618 lần số đứng trước và 0.618 lần số đứng sau nó (0.618 là nghịch đảo của 1.618). Tỷ lệ này được biết đến với rất nhiều tên gọi: tỷ lệ vàng, tỷ lệ thần thánh, PHI ... Vậy thì, tại sao tỷ lệ này lại quan trọng đến vậy? Vạn vật dường như có thuộc tính gắn kết với tỷ lệ 1.618, có lẽ vì thế mà nó được coi là một trong những nhân tố cơ bản cấu thành nên các thực thể trong tự nhiên. Nếu chia tổng số ong cái cho tổng số ong đực trong một tổ ong bất kỳ, bạn sẽ có giá trị là 1.618. Nếu lấy khoảng cách từ vai đến móng tay chia cho khoảng cách giữa cùi chỏ và móng tay thì bạn cũng có được giá trị 1.618. Tính xác thực của các ví dụ trên có thể từ từ kiểm chứng nhưng chúng ta hãy cùng xem “tỷ lệ vàng” có ứng dụng gì trong tài chính.

Khi sử dụng phân tích kỹ thuật, “tỷ lệ vàng” thường được diễn giải theo 3 giá trị phần trăm: 38.2%, 50%, và 61.8%. Nhiều tỷ lệ khác có thể được sử dụng khi cần thiết, như 23.6%, 161.8%, 423%... Có 4 phương pháp chính trong việc áp dụng dãy Fibonacci trong tài chính:

- Fibonacci Retracements
- Fibonacci Arcs
- Fibonacci Fans
- Fibonacci Time zones.

Fibonacci Arcs (FA): được thiết lập đầu tiên bằng cách vẽ đường thẳng kết nối 2 điểm có mức giá cao nhất và thấp nhất của giai đoạn phân tích. 3 đường cong sau đó được vẽ với tâm nằm trên điểm có mức giá cao nhất và có khoảng cách bằng 38.2%, 50.0%, 61.8% độ dài đường thẳng thiết lập FA dùng để dự đoán mức hỗ trợ và kháng cự khi

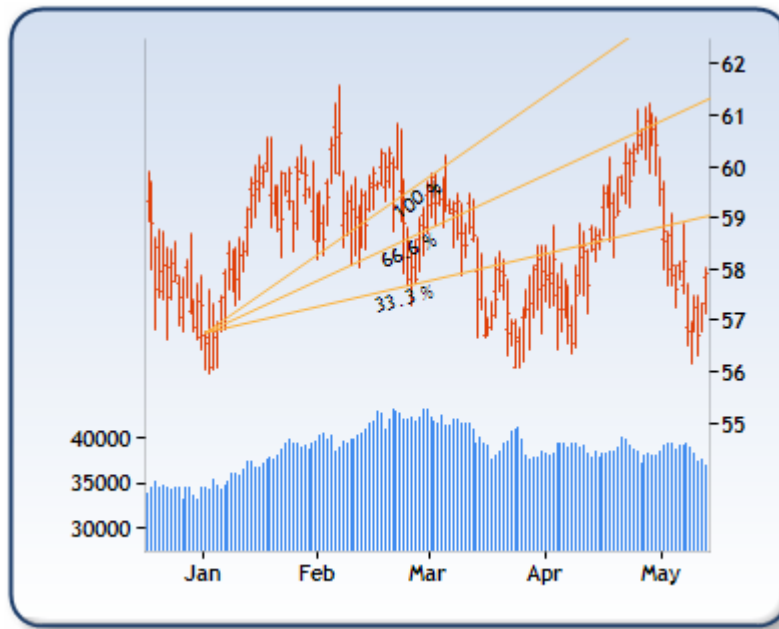
đồ thị giá tiếp cận với đường cong. Một kỹ thuật phổ biến là theo dõi cả hai đường FA, FF (Fibonacci Fans) và dự đoán mức hỗ trợ/kháng cự tại điểm giao giữa đồ thị giá và đường FA/FF.



Hình 49 : Biểu đồ Fibonacci Arcs

Lưu ý rằng đồ thị giá cắt đường FA tại điểm nào còn tùy thuộc vào kích cỡ của đồ thị, nói cách khác đường FA được vẽ lên đồ thị nên nó có mối tương quan với kích cỡ cân đối của đồ thị trên màn hình vi tính hoặc trên giấy. Đồ thị giá của Đồng Bảng Anh mô tả cách mà đường FA tìm ra các điểm hỗ trợ và kháng cự (điểm A, B, C).

Fibonacci Fan (FF): được vẽ bằng cách kết nối hai điểm giá cao nhất và thấp nhất của giai đoạn phân tích. Sau đó một đường thẳng đứng “vô hình” sẽ được vẽ qua điểm giá cao nhất. Tiếp theo đó 3 đường chéo sẽ được vẽ từ điểm giá thấp nhất cắt đường thẳng đứng “vô hình” tại 3 mức 38.2%, 50.0%, 61.8%. Đồ thị sau thấy các ngưỡng hỗ trợ/kháng cự trên đường FF:

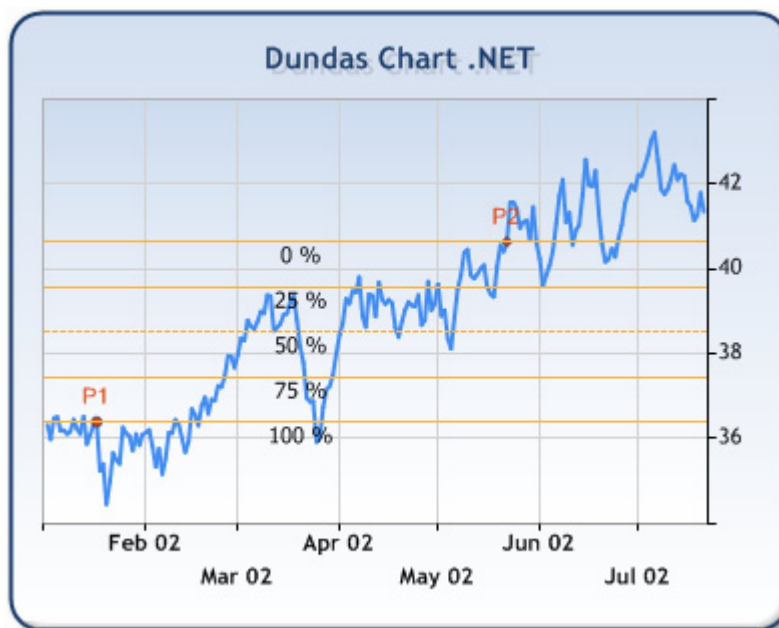


Hình 50 : Biểu đồ Fibonacci Fans

Bạn có thể thấy khi đồ thị giá gặp đường FF cao nhất (điểm A), đồ thị giá không thể vượt qua đường FF trong nhiều ngày. Khi giá vừa vượt qua đường FF, nó liền rút nhanh chóng đến điểm đáy trên đường FF thứ 3 (điểm B và C) trước khi tìm được ngưỡng hỗ trợ. Cũng lưu ý rằng khi giá di chuyển qua điểm đáy (điểm C), nó di chuyển một mạch tới điểm cao nhất (điểm D) trên đường FF thứ 1 và cũng là điểm kháng cự, sau đó rơi xuống điểm giữa trên đường FF thứ hai (điểm E) trước khi đổi chiều đi lên.

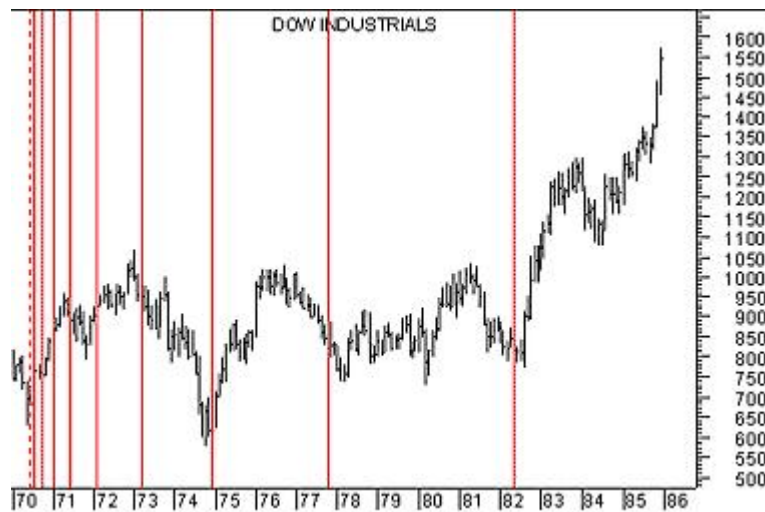
Fibonacci Retracements (FR) được xác định trước tiên bằng cách vẽ đường thẳng nối kết giữa hai điểm giá cao nhất và thấp nhất của đồ thị giá trong giai đoạn phân tích. Một loạt 9 đường nằm ngang sau đó được vẽ lên tại các mức Fibonacci 0.0%, 23.6%, 38.2%, 50%, 61.8%, 100%, 161.8%, 261.8%, và 423.6% tương ứng với chiều cao tính từ điểm giá cao nhất đến thấp nhất (một số đường có thể không được vẽ ra khi nằm

ngoài quy mô phân tích của đồ thị) Sau mỗi giai đoạn biến động giá chính (có thể lên hoặc xuống), giá thường có xu hướng đảo ngược xu hướng (toàn bộ hoặc một phần). Khi giá đảo chiều, các ngưỡng hỗ trợ hoặc kháng cự mới thường nằm trên hoặc gần đường FR (xem đồ thị - ngưỡng hỗ trợ và kháng cự xuất hiện tại đường Fibonacci 23.6% và 38.2%)



Hình 51 : Biểu đồ Fibonacci Retracements

Fibonacci Time Zones bao gồm một loạt các đường thẳng đứng. Sắp xếp theo trật tự của dãy Fibonacci 1, 2, 3, 5, 8, 13, 21, 34, ... Cũng như các đường khác, diễn biến thay đổi của giá thường có mức hỗ trợ/kháng cự nằm gần hoặc trên các đường thẳng đứng này.



Hình 1 : Biểu đồ Fibonacci Time Zones

Các kết quả nghiên cứu về Fibonacci này không có ý định làm kim chỉ nam cho việc xác định thời gian xâm nhập hoặc thoát ra khỏi thị trường. Tuy nhiên, nó có thể hữu ích trong việc xác định vùng hỗ trợ và kháng cự. Đa phần các nhà đầu tư sử dụng kết hợp cả 4 phương pháp Fibonacci này để có thể đưa ra các mức dự đoán chính xác hơn. Một nhà phân tích có thể chỉ sử dụng Fibonacci Arcs và các điểm giao tại vùng hỗ trợ hoặc kháng cự. Nhiều người khác kết hợp các nghiên cứu về Fibonacci với các dạng thức phân tích kỹ thuật khác như “lý thuyết sóng Elliott” để dự đoán mức độ đảo ngược xu hướng sau mỗi bước sóng khác nhau. Trong phạm vi khóa luận, chỉ xin dừng lại ở mức độ ứng dụng cơ bản của dãy Fibonacci, hy vọng có thể cung cấp một vài thông tin hữu ích.

Kết hợp Fibonacci và sóng Elliott dự đoán xu hướng thị trường

Lý thuyết sóng Elliott

Lý thuyết sóng Elliott được đặt tên theo tên tác giả Ralph Nelson Elliott phát triển từ lý thuyết Dow và bằng những quan sát được tìm thấy một cách tự nhiên. Elliott đã kết luận rằng sự chuyển động của thị trường chứng khoán có thể được dự đoán dựa trên những quan sát và xác định dựa trên những mẫu sóng lặp lại. Thực vậy, Elliott tin rằng tất cả các hoạt động của con người, không chỉ trong thị trường chứng khoán, đều bị ảnh hưởng bởi một chuỗi các sóng có thể xác định được.

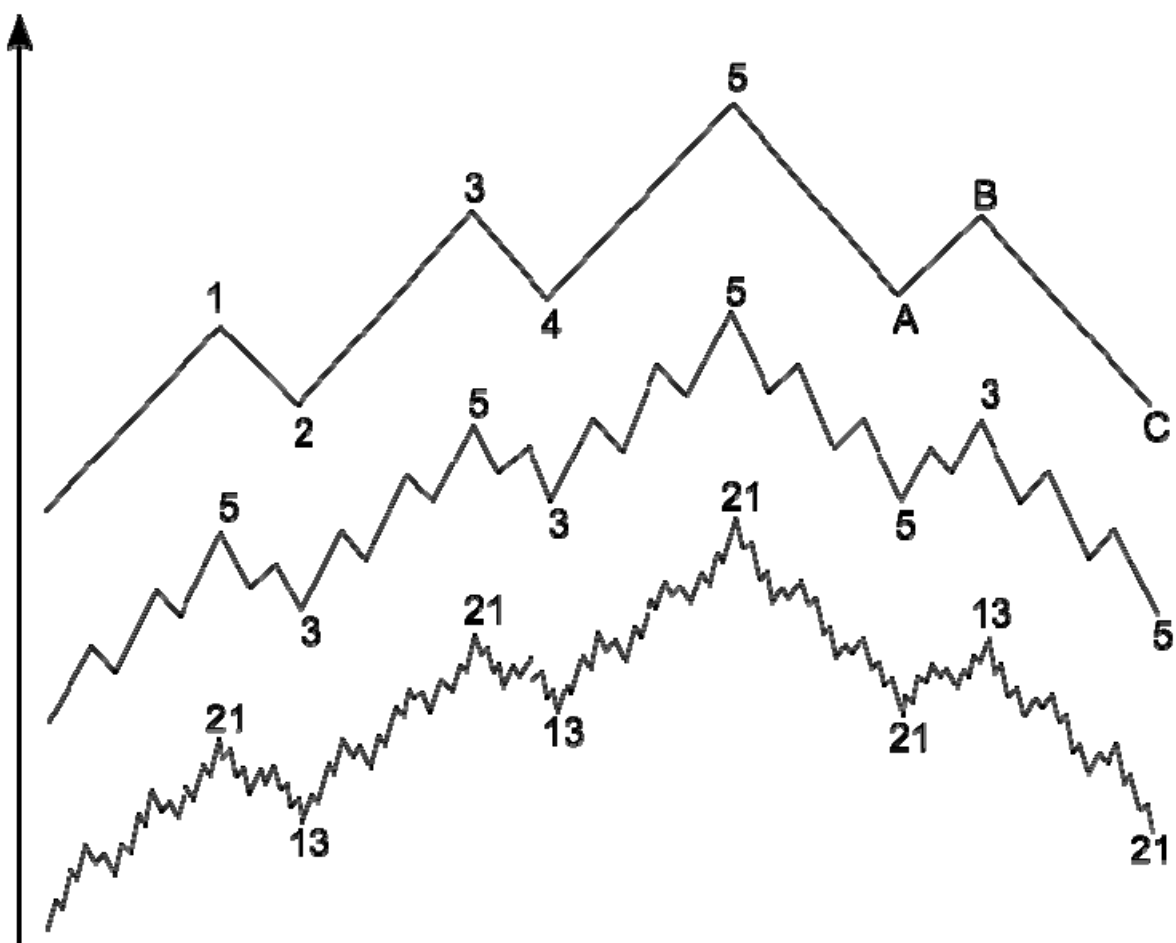
Với sự giúp đỡ của C.J. Collins, ý tưởng của Elliott được chú ý ở Wall Street trong loạt bài được xuất bản trong tạp chí “Financial World” vào năm 1939. Trong suốt những năm thập kỷ 1950 và 1960, những ý tưởng của Elliott được phát triển tiếp bởi Haminton Bolton. Vào 1960, Bolton đã viết cuốn sách “Elliott Wave Principle – A Critical Appraisal”. Đây là công việc hết sức ý nghĩa đầu tiên từ khi Elliott qua đời. Vào năm 1978, Robert Prechter và A.J.Frost đã hợp tác viết cuốn sách “Elliott Wave Principle”.

Theo ông Elliott, sự thay đổi của giá cả sẽ tạo ra những cơn sóng. Trong đó một cơn sóng cơ bản sẽ có 5 cơn sóng “chủ” và 3 cơn sóng điều chỉnh. Trong 5 cơn sóng chủ thì sóng số 1, 3 và số 5 gọi là sóng “chủ và động”, và sóng 2, 4 gọi là sóng “chủ và điều chỉnh”. 2 cơn sóng điều chỉnh được gọi là sóng A,B,C. Trong mỗi một cơn sóng như vậy lại có những cơn sóng nhỏ và cũng tuân theo quy luật của lý thuyết Elliot. Một đợt sóng chủ hoàn chỉnh sẽ có 89 sóng và một đợt sóng điều chỉnh hoàn chỉnh sẽ có 55 sóng.

Tùy theo thời gian độ lớn của sóng sẽ được phân theo thứ tự sau:

- Grant Supercycle: sóng kéo dài nhiều thập kỷ, đôi khi cả thế kỷ
- Supercycle: kéo dài từ vài năm đến vài thập kỷ
- Cycle: kéo dài từ 1 đến vài năm
- Primary: kéo dài từ vài tháng đến vài năm
- Intermediate: kéo dài từ vài tuần đến vài tháng

- Minor: kéo dài trong vài tuần
- Minute: Kéo dài trong vài ngày
- Minuette: Kéo dài trong vài giờ.
- Subminutte: Kéo dài trong vài phút.



Hình 52 : Biểu đồ sóng Elliott và các số Fibonacci

Dãy số Fibonacci

Dãy số do nhà toán học người Ý Leonardo Fibonacci (1175-1250) phát minh ra. Bắt đầu là số 0 và số 1, sau đó là những số kế tiếp được tạo ra bằng cách cộng 2 số đứng

trước. Ví dụ $1 = 1+0$, $5 = 3+2$, $34=21+13$. Điều kỳ diệu hơn là cứ lấy số lớn chia cho số nhỏ hơn một bậc, ví dụ $89/55$ ta sẽ được 1.618; lấy số nhỏ chia cho số lớn hơn 1 bậc, ví dụ $21/34$ ta sẽ được 0.6180, lấy số nhỏ chia cho số lớn hơn 2 bậc, ví dụ $13/34$ ta sẽ được 0.382.

Tất cả các con số thuộc dãy số Fibonacci 0, 1, 1, 2, 3, 5, 8 ...và các con số 0.618 và 0.382, trong đó đặc biệt nhất là con số được mệnh danh là “tỷ lệ vàng” 1.618 - xuất hiện rất nhiều trong tự nhiên, trong cơ thể con người, trong vũ trụ, trong kiến trúc, xây dựng.

Quan trọng hơn đối với chúng ta những nhà đầu tư, kinh doanh chứng khoán, những con số “đầy ma lực trên” bên trên xuất hiện ngay trong thị trường tài chính, nhất là những biến động về giá cả.

Ralph Nelson Elliott khẳng định rằng ông nghiên cứu và phát minh ra lý thuyết sóng trước khi biết Fibonacci nhưng những con số trùng hợp đến kỳ lạ: 5 sóng chủ, 3 sóng điều chỉnh, 89 sóng chủ, 55 sóng điều chỉnh cũng như tỷ lệ giá của các con sóng luôn ở chung quanh các tỷ lệ vàng 0.618, 1.618, 0.328. Do đó có một giả thuyết khác cho rằng ông Elliott đã ứng dụng những con số Fibonacci vào lý thuyết của mình.

Dãy số Fibonacci	Số lớn/chia cho số nhỏ hơn 1 bậc	Số nhỏ/chia cho số lớn hơn 1 bậc	Số nhỏ/chia cho số lớn hơn 2 bậc
0			
1			
1			
2			
3			
5			

8			
13	1.6250	0.6154	0.3810
21	1.6154	0.6190	0.3824
34	1.6190	0.6176	0.3818
55	1.6176	0.6182	0.3820
89	1.6182	0.6180	0.3819
144	1.6180	0.6181	0.3820
233	1.6181	0.6180	0.3820
377	1.6180	0.6180	0.3820
610	1.6180	0.6180	0.3820
987	1.6180	0.6180	

Bảng 16 : Bảng thể hiện tỉ lệ trong chuỗi Fibonacci

Một dãy sóng 5-3 điển hình trong thị trường tăng trưởng “bò húc” (bull market). Dưới đây là phân tích một con sóng 5-3 điển hình của thị trường trong giai đoạn tăng trưởng - “bò húc”. Cũng con sóng 5-3 này trong thị trường suy thoái – “gấu ngủ” (bear market) sẽ được vẽ hoàn toàn ngược lại.

Sóng chủ số 1: Đợt sóng đầu tiên này là có điểm xuất phát từ thị trường con gấu (suy thoái), do đó sóng 1 ít khi được nhận biết ngay từ đầu. Lúc này thông tin cơ bản về các công ty niêm ước vẫn đang là thông tin tiêu cực. Chiều hướng của thị trường trước khi sóng 1 xảy ra chủ yếu vẫn là thị trường suy thoái. Những nhà phân tích cơ bản vẫn đang tiếp tục điều chỉnh thu nhập kỳ vọng thấp xuống so với dự kiến. Khối lượng giao dịch có tăng chút ít theo chiều hướng giá tăng. Tuy vậy việc tăng này là không đáng kể. Do đó nhiều nhà phân tích kỹ thuật không nhận ra sự có mặt của đợt sóng số 1 này.

Sóng chủ số 2: Sóng chủ 2 sẽ điều chỉnh sóng 1, nhưng điểm thấp nhất của sóng 2 không bao giờ vượt qua điểm xuất phát đầu tiên của sóng 1. Tin tức dành cho thị trường vẫn chưa khả quan. Thị trường đi xuống ở cuối sóng 2 để thực hiện việc “kiểm

tra” độ thấp của thị trường. Những người theo phái con gấu vẫn đang tin rằng thị trường con gấu vẫn đang ngự trị. Khối lượng giao dịch sẽ ít hơn đợt sóng 1. Giá sẽ được điều chỉnh giảm và thường nằm trong khoảng 0.382 đến 0.618 của mức cao nhất của sóng 1.

Sóng chủ số 3: Thông thường đây là sóng lớn nhất và mạnh mẽ nhất của xu hướng lên giá. Ngay đầu sóng 3, thị trường vẫn còn nhận những thông tin tiêu cực vì vậy có nhiều nhà kinh doanh không kịp chuẩn bị để mua vào. Khi sóng 3 đang ở lưng chừng, thị trường bắt đầu nhận những thông tin cơ bản tích cực và những nhà phân tích cơ bản bắt đầu điều chỉnh thu nhập kỳ vọng. Mặc dù có những đợt điều chỉnh nho nhỏ trong lòng của sóng 3, giá của sóng 3 tăng lên với tốc độ khá nhanh. Điểm cao nhất của sóng 3 thường cao hơn điểm cao nhất của sóng 1 với tỷ lệ 1,618:1

Sóng chủ số 4: Đây thật sự là một con sóng điều chỉnh. Giá có khuynh hướng đi xuống và đôi khi có thể là hình răng cưa kéo dài. Sóng 4 thường sẽ điều chỉnh sóng 3 với mức 0.382 – 0.618 của sóng 3. Khối lượng giao dịch của sóng 4 thấp hơn của sóng 3. Đây là thời điểm để mua vào nếu như nhà kinh doanh nhận biết được tiềm năng tiếp diễn liên sau đó của con sóng 5. Tuy vậy việc nhận biết điểm dừng của sóng 4 là một trong những khó khăn của các nhà phân tích kỹ thuật trường phái sóng Elliott.

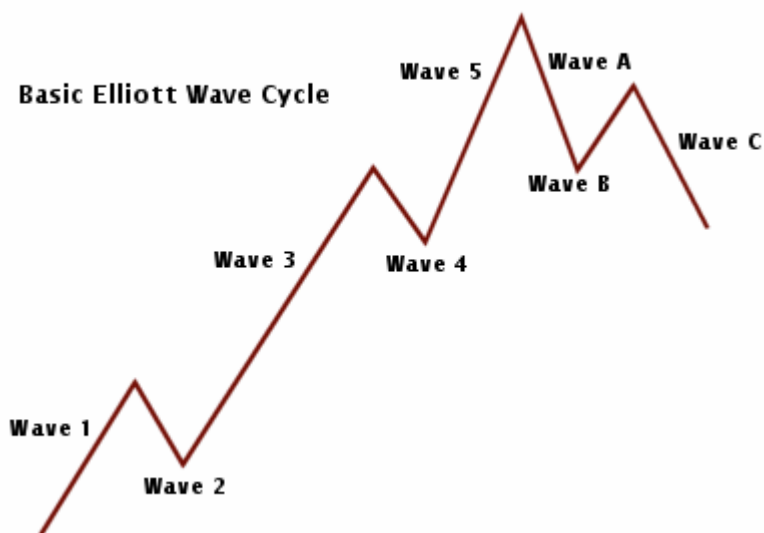
Sóng chủ số 5: Đây là đợt sóng cuối cùng của 5 con sóng “chủ”. Thông tin tích cực tràn lan khắp thị trường và ai cũng tin rằng thị trường đang ở trong thế “bò húc”. Khối lượng giao dịch của sóng 5 khá lớn, tuy vậy thông thường vẫn nhỏ hơn sóng 3. Điều đáng nói là những nhà kinh doanh “không chuyên nghiệp” thường mua vào ở những điểm gần cuối sóng 5. Vào cuối con sóng 5, thị trường nhanh chóng chuyển hướng.

Sóng điều chỉnh A: Sóng này bắt đầu cho đợt sóng điều chỉnh A,B,C. Trong thời gian diễn ra sóng A, thông tin cơ bản vẫn đang rất lạc quan. Mặc dù giá xuống, nhưng phần

đông các nhà kinh doanh vẫn cho rằng thị trường đang trong thế bò húc. Khối lượng giao dịch tăng trưởng khá đều đặn theo con sóng A.

Sóng điều chỉnh B: Giá tăng trở lại và với mức cao hơn so với điểm cuối sóng A. Sóng B được xem là điểm kéo dài của thị trường bò húc. Đối với những người theo trường phái phân tích kỹ thuật cổ điển, điểm B chính là vai phải của đồ thị “Đầu và Vai” ngược. Khối lượng giao dịch của sóng B thường thấp hơn của sóng A. Vào lúc này, những thông tin cơ bản của các công ty không có những điểm tích cực mới, thế nhưng cũng chưa chuyển hẳn qua tiêu cực.

Sóng điều chỉnh C: Giá có khuynh hướng giảm nhanh hơn các đợt sóng trước. Khối lượng giao dịch tăng. Hầu như tất cả mọi nhà kinh doanh, đầu tư đều nhận thấy rõ sự ngụt trị của “gấu ngủ” trên thị trường, chậm nhất là trong đợt sóng nhỏ thứ 3 của sóng C. Sóng C thường lớn như sóng A. Điểm thấp nhất của sóng C ít nhất bằng điểm thấp nhất của sóng A nhân với 1.618.



Hình 53 : Sóng Elliott cơ bản

Tóm gọn về Fibonacci và sóng Elliott: Tuy Lý thuyết sóng Elliott kết hợp với dãy số Fibonacci bị một số chỉ trích phê bình, nó vẫn được rất nhiều nhà nghiên cứu và kinh doanh các sản phẩm tài chính cổ vũ và sử dụng trong việc phân tích giá. Và càng ngày nó càng trở nên phổ biến vì sự chính xác đôi khi đến bất ngờ của nó. Trong hội thảo đầu tư tài chính châu Á, ngày 26/7/2007 tại Thành Phố Hồ Chí Minh, ông Joe DiNapoli, một chuyên gia về kỹ thuật Fibonacci đã chứng minh rằng trong năm vừa rồi, có hai con sóng của VNIndex đã theo đúng các con số Fibonacci [13]. Dĩ nhiên chúng ta có thể tìm ra thêm nhiều chứng cứ xác thực của VNIndex cho “lý thuyết sóng Elliott kết hợp với Fibonacci”. Điều đó cũng giống như nhiều dụng cụ phân tích kỹ thuật khác: khá chính xác khi chứng minh quá khứ, nhưng chính xác “vừa phải” khi dự đoán tương lai. Do đó khi sử dụng lý thuyết trên cũng như bất cứ phương pháp/dụng cụ phân tích kỹ thuật nào khác, chúng ta phải hết sức thận trọng, sáng suốt và quan trọng hơn hết là chuẩn bị chấp nhận một mức độ rủi ro nhất định (Theo [13]).

