

MỘT GIẢI PHÁP MỚI VỀ PHÂN TÍCH THỊ TRƯỜNG CHỨNG KHOÁN VIỆT NAM

CÙ THU THUYẾT¹, ĐỖ VĂN THÀNH²

¹*Khoa Hệ thống Thông tin Kinh tế - Học viện Tài chính*

²*Bộ Kế hoạch và Đầu tư*

Abstract. The aim of this paper is to propose a new solution for analyzing Vietnamese stock market based on the approach of data mining. By finding out and analyzing association rules mined in Vietnamese stock data set announced officially by Hanoi Securities Trading Center, some hidden relationships between the increasing or decreasing of some stock prices internally, and between some stock prices and HASTC index are indicated. The results show that these association rules reflect relatively real developments of stock market in Vietnam over the last few years.

Tóm tắt. Bài báo đề xuất giải pháp phân tích thị trường chứng khoán Việt Nam theo cách tiếp cận của khai phá dữ liệu (data mining). Cụ thể là bằng cách phát hiện và phân tích các luật kết hợp phát hiện được từ tập dữ liệu chứng khoán do Trung tâm giao dịch chứng khoán Hà Nội công bố chính thức để phát hiện ra các mối quan hệ tiềm ẩn về tăng hoặc giảm giá giữa các mã cổ phiếu với nhau và giữa các mã cổ phiếu với Haste Index. Những luật kết hợp được phát hiện đã phản ánh khá sát thực tế diễn biến của thị trường chứng khoán Việt Nam thời gian qua.

1. GIỚI THIỆU VẤN ĐỀ

Thực tiễn phát triển của thị trường chứng khoán Việt Nam thời gian qua cho thấy thị trường này phát triển chưa được vững chắc, còn tiềm ẩn nhiều yếu tố rủi ro và trong nhiều trường hợp, sự biến động của thị trường không tuân theo những quy luật kinh tế thông thường và rất khó giải thích sự biến động của nó bằng các lý thuyết kinh tế, tài chính tiền tệ đã biết.

Để thị trường chứng khoán Việt Nam thực sự trở thành kênh thu hút vốn đầu tư quan trọng cho phát triển kinh tế - xã hội, điều cần thiết là phải có những công cụ phân tích kỹ thuật và dự báo các yếu tố ảnh hưởng đến sự phát triển của thị trường này ở nhiều khía cạnh nhằm giúp các cơ quan quản lý nhà nước xây dựng được cơ chế, chính sách phù hợp để duy trì sự ổn định và phát triển thị trường và giúp các nhà đầu tư có thêm cơ sở khoa học để tin tưởng và quyết định đầu tư hiệu quả.

Hiện tại, ở nước ta một số công cụ hỗ trợ phân tích thị trường chứng khoán đã được phát triển và ứng dụng rộng rãi [6, 9 – 12]. Tuy nhiên, có thể nói cho đến thời điểm này trong thị trường tài chính, tiền tệ quốc tế nói chung, ở nước ta nói riêng, các công cụ đã được sử dụng để phân tích và dự báo về thị trường chứng khoán [2] đều không hỗ trợ hiệu quả cho việc trả lời các câu hỏi: sự tăng hay giảm về giá cổ phiếu của những nhóm công ty tham gia chứng khoán nào có tác động đến sự tăng, giảm của các chỉ số chứng khoán? Có tồn tại mối

quan hệ giữa các chỉ số chứng khoán của các Trung tâm giao dịch chứng khoán khác nhau cả trong và ngoài nước hay không? Có tồn tại mối quan hệ về sự biến động về giá cổ phiếu giữa một số nhóm công ty tham gia vào thị trường chứng khoán không và bằng cách nào có thể xác định được những nhóm công ty có mối quan hệ đó,...

Mục đích của bài báo này là trả lời một phần những câu hỏi nêu trên cho trường hợp thị trường chứng khoán Việt Nam bằng cách sử dụng cách tiếp cận của khai phá dữ liệu (data mining) để phát hiện ra các mẫu, các quan hệ tiềm ẩn trong tập dữ liệu thực tế về các công ty tham gia thị trường chứng khoán và về chỉ số chứng khoán Hasc.

Bài báo đã sử dụng kỹ thuật phát hiện luật kết hợp, đó là một trong những hướng nghiên cứu và ứng dụng rất quan trọng hiện nay của khai phá dữ liệu.

Sau phần giới thiệu, Mục 2 sẽ trình bày một số phân tích ban đầu tập dữ liệu chứng khoán được thu thập từ Trung tâm giao dịch chứng khoán Hà Nội, cách tiếp cận lựa chọn kỹ thuật khai phá dữ liệu từ tập dữ liệu được thu thập và phần mềm công cụ hỗ trợ kỹ thuật đó. Mục 3 sẽ thực hiện phát hiện luật kết hợp trên tập dữ liệu và phân tích các luật kết hợp vừa phát hiện được. Một số kết luận về phương pháp khai phá dữ liệu được thực hiện trong bài báo.

2. LỰA CHỌN KỸ THUẬT KHAI PHÁ DỮ LIỆU

2.1. Thu thập và phân tích ban đầu dữ liệu chứng khoán

Dữ liệu chứng khoán được thu thập từ các thông tin, số liệu được công bố hàng ngày của Trung tâm giao dịch chứng khoán Hà Nội và trên [10, 12], sau đó được tổ chức thống nhất trên bảng tính Excel như sau.

NGÀY		09-th3-07		12-th3-07		13-th3-07		14-th3-07		15-th3-07		16-th3-07	
Hasc Index		454,81		454,75		446,95		431,77		421,33		448,39	
PHIÊN		327		328		329		330		331		332	
STT	MÃ	Gia	KL	Gia	KL	Gia	KL	Gia	KL	Gia	KL	Gia	KL
1	ACB	288,8	70200	278,7	145900	271,4	121900	261,3	124800	254,5	228200	277,6	99500
2	BBS	47,3	163800	44,2	123900	44,4	85700	42,3	44400	38,6	27300	37,6	41000
3	BCC	51	462000	52,3	459700	53	385100	51,8	263400	47,3	224100	51,4	254300
4	BHV	15,4	0	15,4	0	15,4	0	15,4	0	15,4	0	15,4	0
5	BMI	207,5	55300	211,2	30500	197,9	18700	181,3	75600	185,9	30600	203,5	17800
6	BTS	49,6	124900	52,2	134800	52,2	107000	51,4	56100	47,5	196200	49,7	45300
7	BVS	437,6	36600	435,4	22700	420,4	37600	384,1	86900	373,3	44900	409,1	21900
8	CIC	79,6	21900	73,3	13900	70,6	7400	67,2	6700	62,3	19000	68,5	700
9	CID	61,7	12500	67,6	10100	68,3	4800	64,4	1900	57,7	1800	52,2	4400
10	CJC	69,7	9800	76,6	7600	77	3600	73,1	3000	70,2	700	62,4	1900
11	CMC	62,9	2900	69,1	4100	75,9	18200	71,8	1100	68,7	400	64,5	2600
12	CTB	61,6	20600	62,6	3900	56,4	4100	50,8	3600	45,8	3000	42	12400
13	CTN	90,2	8300	88,6	13500	86,9	15300	86,8	34600	83,8	16800	85,4	20800
14	DAC	39,7	3800	43,6	9600	47,9	40300	48,3	20400	45,2	5400	40,8	24200
15	DAE	62,9	1000	69	13100	69,4	4500	62,5	1600	56,3	300	53,9	5600
16	DHI	76,5	22000	79,7	15200	71,8	7900	64,7	13800	58,1	3500	56,8	15400
17	DTC	57,3	1000	62,7	4900	62,6	2100	60,7	300	54,7	200	60,1	2600
18	EBS	150,7	12700	158	9800	152,9	6500	141,8	5200	129,4	8000	142,3	1400
19	GHA	73	20600	72,5	10900	67,8	5300	61,7	3500	55,6	2200	52,5	12000
20	HAI	88,3	127600	81,2	119600	86	46200	83	25000	77,3	17900	81,6	39600
21	HJS	62,7	28500	59,7	38200	56,9	44000	54,2	26800	49,2	8200	45,8	86400
22	HLY	65,5	2100	71,1	1700	70,9	2200	68,9	100	62,2	1600	56	1600
23	HNH	63	184400	67,4	315100	68,6	223000	66,7	81900	60,5	69700	64,2	137100
24	HPC	160,1	66500	160,9	64700	159,7	59500	157,7	35400	147,2	50400	160,3	42800
25	HPS	36,6	28800	40,2	45400	44,2	178800	45,7	83700	41,4	19800	40,1	47100
26	HSC	72,1	0	79,3	100	79,3	500	71,4	2000	64,3	100	57,9	100
27	HTP	58,9	18800	64,4	22000	59,5	10200	54,2	12600	48,8	1900	49,5	31800
28	ICF	54,5	67900	59,9	182700	57,4	163100	51,7	56100	46,6	53400	47	170000

Hình 1. Tổ chức dữ liệu về các phiên giao dịch trên Excel

Tập dữ liệu này lưu trữ thông tin về các phiên giao dịch chứng khoán từ ngày 14/7/2005 đến ngày 12/9/2007. Tổng số có 88 mã cổ phiếu và 456 phiên giao dịch. Các phiên giao dịch được lưu trữ liên tục theo thời gian và gồm các thông tin: số hiệu phiên giao dịch, Haste Index, mã chứng khoán, giá bình quân cuối ngày và khối lượng giao dịch của mỗi mã chứng khoán.

Phân tích ban đầu tập dữ liệu cho thấy có những phiên giao dịch giá cổ phiếu của hầu hết các mã chứng khoán không biến động so với phiên trước đó, nhưng cũng có nhiều phiên có rất nhiều mã cổ phiếu đồng loạt tăng giá hoặc đồng loạt giảm giá. Số mã cổ phiếu đồng loạt tăng giá nhiều nhất trong một phiên là 78 mã và số mã cổ phiếu biến động giảm giá nhiều nhất trong một phiên là 46 mã. Phân tích ban đầu tập dữ liệu cũng cho thấy sự biến động tăng, giảm giá của nhiều mã cổ phiếu cũng tác động đến sự tăng, giảm của Haste Index.

2.2. Lựa chọn kỹ thuật khai phá dữ liệu và giới hạn phạm vi bài toán

Nếu coi mỗi mã cổ phiếu, cũng như Haste Index như là một mục dữ liệu. Thế thì tập dữ liệu được mô tả như trên có thể được xem là một cơ sở dữ liệu (CSDL) định lượng [3, 4], đặc biệt nếu ta chỉ quan tâm đến giá cổ phiếu là tăng hoặc giảm hay không thì ta sẽ nhận được một CSDL tác vụ [1, 5] về các phiên giao dịch chứng khoán. Việc tìm câu trả lời cho những câu hỏi đặt ra thực chất là đi tìm mối quan hệ giữa các mục dữ liệu (hay thuộc tính) trong CSDL tác vụ (hay CSDL định lượng) đã cho và vì thế có thể ứng dụng kỹ thuật Phát hiện luật kết hợp từ CSDL để giải quyết vấn đề đặt ra.

Về bản chất luật kết hợp diễn tả mối quan hệ giữa các nhóm mục dữ liệu (hoặc thuộc tính) trong các CSDL tác vụ (hoặc CSDL định lượng). Phát hiện luật kết hợp được giới thiệu lần đầu vào năm 1993, được phát triển mạnh từ năm 1996 và hiện đã trở thành một trong những hướng nghiên cứu, ứng dụng quan trọng của khai phá dữ liệu [1, 5, 7]. Vấn đề phát hiện luật kết hợp đã được ứng dụng rất thành công trong nhiều lĩnh vực kinh tế - xã hội khác nhau như: thương mại, y tế, tin sinh học, khai phá văn bản, khai phá web,... Bài báo này có thể được xem là nghiên cứu ban đầu về việc ứng dụng Phát hiện luật kết hợp trong lĩnh vực tài chính - chứng khoán.

Trong vấn đề Phát hiện luật kết hợp, trước hết người ta tập trung nghiên cứu đề xuất những thuật toán nhằm phát hiện luật kết hợp từ các CSDL tác vụ và sau đó phát triển các thuật toán này để có thể phát hiện được các luật kết hợp từ CSDL định lượng (các thuộc tính nhận giá trị số hoặc phân loại) [3, 4]. Bài báo này mới chỉ tập trung vào phát hiện luật kết hợp từ CSDL tác vụ.

Dạng tổng quát của luật kết hợp được phát hiện từ các CSDL tác vụ có dạng $X \rightarrow Y (\alpha\%, \beta\%)$, ở đây X, Y là tập gồm một số các mục dữ liệu, $X \cap Y = \emptyset$, $\alpha\%, \beta\%$ biểu thị tương ứng cho mức độ hỗ trợ và tin cậy của luật. X được gọi là tiền đề, Y được gọi là hệ quả.

Để phát hiện được các luật kết hợp cần phải tiến hành 2 giai đoạn liên quan đến độ hỗ trợ và độ tin cậy của luật. Giai đoạn thứ nhất nhằm tìm ra các tập mục dữ liệu có độ hỗ trợ lớn hơn hoặc bằng một số minSup cho trước được gọi là độ hỗ trợ cực tiểu và tập đó được gọi là tập phổ biến. Giai đoạn thứ hai tìm các luật kết hợp từ các tập phổ biến có độ tin cậy lớn hơn hoặc bằng một số minConf cho trước và được gọi là độ tin cậy cực tiểu [1, 5]. Trong quá trình phát hiện luật kết hợp thì giai đoạn đầu là phức tạp và tốn nhiều chi phí nhất [8].

Trong trường hợp CSDL tác vụ chứng khoán, mỗi mục dữ liệu sẽ chính là một mã chứng khoán, và luật $X \rightarrow Y (\alpha\%, \beta\%)$ có nghĩa là $\alpha\%$ của tổng số các phiên giao dịch có các mã cổ phiếu trong $X \cup Y$ tăng (giảm) giá, và $\beta\%$ các phiên giao dịch có các mã cổ phiếu trong X tăng (giảm) giá cũng là các phiên giao dịch có các mã cổ phiếu trong Y tăng (hoặc giảm) giá, hay nói cách khác $\beta\%$ các phiên giao dịch có X là tiền đề dẫn đến hệ quả Y .

- Việc tìm ra các luật dạng $X \rightarrow Y$ từ CSDL tác vụ chứng khoán sẽ cho phép tăng cường việc kết xuất được thông tin, phát hiện được những tri thức mới từ dữ liệu, chẳng hạn:

- Giúp nhận diện được tất cả các luật có các mã chứng khoán trong Y như là hệ quả. Các luật này giúp những nhà quản lý và đầu tư nhận diện được việc tăng (hoặc giảm) giá của các mã cổ phiếu trong Y do các mã cổ phiếu nào chi phối và chi phối ở mức nào.

- Nhận diện được tất cả các luật có X là tiền đề. Các luật này giúp xác định được những mã cổ phiếu nào sẽ bị ảnh hưởng khi các mã cổ phiếu trong X tăng (giảm) giá và ảnh hưởng ở mức nào.

- Nhận diện được tất cả các luật có các mã cổ phiếu trong X nằm ở phần tiền đề, các mã cổ phiếu trong Y nằm ở hệ quả. Các luật này giúp các nhà quản lý và đầu tư chứng khoán biết được có thể làm cho một số mã cổ phiếu khác cùng tăng (hoặc giảm) giá với các mã cổ phiếu trong X trong khi vẫn tôn trọng mong muốn để cho các mã cổ phiếu trong Y tăng (hoặc giảm) giá.

- Tìm k luật tốt nhất đối với một số mã chứng khoán xác định nào đó nằm ở phần hệ quả, các luật này rất hữu ích giúp các đầu tư chứng khoán nhận biết dấu hiệu tăng (hoặc giảm) giá của các mã chứng khoán đó thông qua sự tăng (giảm giá) của các mã chứng khoán khác, để từ đó có quyết định đầu tư mua vào hay bán ra cổ phiếu mà mình đang nắm giữ,...

- Và còn rất nhiều phân tích khác nữa kiểu như vậy.

2.3. Công cụ hỗ trợ phát hiện luật kết hợp

Hiện có khá nhiều nhóm nghiên cứu về luật kết hợp theo các cách tiếp cận xây dựng thuật toán tìm tập phổ biến khác nhau và do đó cũng có nhiều phần mềm tìm luật kết hợp từ các CSDL tác vụ khác nhau. Hầu hết các phần mềm đó chưa được trở thành thương mại. Trong bài báo này sẽ sử dụng phần mềm CBA (Classification Based on Associations) do trường Đại học Quốc gia Singapo phát triển [13]. Trong phần mềm CBA đã sử dụng thuật toán tìm tập phổ biến Apriori do R. Agrawal và cộng sự, đó là những người đầu tiên mở hướng nghiên cứu và ứng dụng về luật kết hợp, đề xuất [1]. Thuật toán Apriori được đánh giá ở mức trung bình về độ phức tạp tính toán [8]. CBA có thể hỗ trợ để phát hiện luật kết hợp từ CSDL quan hệ và CSDL tác vụ; hỗ trợ phát hiện luật kết hợp với nhiều độ hỗ trợ cực tiểu.

Một luật kết hợp tìm được qua phần mềm CBA có dạng [13]:

$\mathbf{A} = \mathbf{Y} \rightarrow \mathbf{B} = \mathbf{Y}(\text{Cover}\%, \text{Conf}\%, \text{CoverCount}, \text{SupCount}, \text{Sup}\%)$, ở đây \mathbf{A}, \mathbf{B} là tập các mục dữ liệu, $\mathbf{A} = \mathbf{Y}$ có nghĩa là tất cả các mục dữ liệu trong \mathbf{A} xảy ra;

$\text{Cover}\%, \text{Conf}\%, \text{CoverCount}, \text{SupCount}, \text{Sup}\%$ là 5 thông số của luật, trong đó:

- $\text{Cover}\%$ cho biết số phần trăm các phiên giao dịch so với tổng số có sự tăng (hoặc giảm) giá của các mã cổ phiếu trong A .

- $\text{Conf}\%$, gọi là độ tin cậy của luật được tính bằng công thức $\text{SupCount}/\text{CoverCount}$ với SupCount và CoverCount là các thông số được định nghĩa ở dưới.

- CoverCount cho biết số phiên giao dịch có sự xuất hiện tăng (hoặc giảm) giá của các mã cổ phiếu trong A .

- *SupCount* cho biết số phiên giao dịch có sự xuất hiện tăng (hoặc giảm) giá của các mã cổ phiếu trong cả 2 tập *A, B*.

- *Sup%*, gọi là độ hỗ trợ của luật, đó là số phần trăm các phiên giao dịch so với tổng số ở đó các mã cổ phiếu trong cả 2 tập *A* và *B* đều tăng (hoặc giảm) giá.

Về bản chất luật này không khác biệt so với luật kết hợp tổng quát được phát hiện từ CSDL tác vụ bởi trong 5 thông số trên, 2 thông số *Sup%*, *Conf%* tương ứng là độ hỗ trợ và độ tin cậy của luật kết hợp đã được giới thiệu trước đó, 3 thông số còn lại thực ra chỉ nhằm để làm rõ hơn về 2 thông số vừa nêu.

Các luật kết hợp tìm được bởi phần mềm CBA có thể được trình bày dưới nhiều hình thức như liệt kê tuần tự, thể hiện dưới dạng cây, có thể được lưu ở dạng tệp văn bản,..., do đó dễ dàng đọc và sử dụng các luật được phát hiện.

3. PHÁT HIỆN LUẬT KẾT HỢP CHỨNG KHOÁN

3.1. Tập dữ liệu để phát hiện luật kết hợp chứng khoán

Do có nhiều phiên giao dịch có rất ít biến động về tăng (giảm) giá và có số lượng cổ phiếu giao dịch thấp, nên để độ hỗ trợ của luật cao rất cần loại bỏ các phiên giao dịch này khỏi cơ sở dữ liệu chứng khoán. Theo cách như vậy, số lượng các phiên giao dịch được giữ lại để nghiên cứu đã giảm từ 456 xuống còn 186 phiên.

BBS, BHV, BMI, DHI, HSC, ILC, NTP, TBC, TKU, VC2, VMC, VNR
 ACB, NTP, PTS, VC2, VMC, VNR, VTL, Hastc Index
 ACB, BCC, BMI, BTS, CMC, DAC, NTP, STP, TBC, VC2, VMC, VNR, Hastc Index
 ACB, BCC, BTS, CJC, CMC, HTP, MEC, NLC, NTP, PTS, SAP, SJE, TBC, TKU, VC2, VMC, VNR, Hastc Index
 ACB, BBS, BCC, BMI, BTS, CJC, CMC, HPC, HSC, MEC, NLC, SJE, SSI, TBC, TKU, TLT, TXM, VC2, VMC, VNR, Hastc Ind
 BCC, BMI, BTS, BVS, CJC, HPC, ICF, LTC, MEC, NTP, SJE, SSI, TBC, TKU, TLT, VC2, VMC, VNR, VTV, Hastc Index
 ACB, BCC, BMI, BTS, CJC, CMC, DHI, HPC, ICF, MEC, NLC, NTP, PTS, SSI, TBC, TLT, TXM, VC2, VNR, VTL, VTV, Hast
 BMI, BTS, CJC, CMC, HJS, HPC, ICF, LTC, MEC, PPG, STP, TBC, TXM, VC2, VNR, VTV
 BBS, CMC, GHA, TKU, TXM, VTS, VTV
 BBS, CID, CMC, CTB, HSC, NTP, TKU, TXM, VTV
 ACB, BBS, BCC, BMI, BVS, CMC, EBS, HPC, ICF, PAN, PJC, POT, S99, SSI, TBC, TXM, VC2, VMC, VTL, VTS, VTV, Hast
 ACB, BBS, BCC, BMI, BTS, CJC, CMC, EBS, HLY, HPC, ICF, NLC, NTP, PAN, PJC, POT, PTC, PTS, SSI, TBC, TLT, TPH,
 ACB, BMI, BTS, CMC, CTN, DHI, HAI, HLY, HPC, MEC, NLC, PAN, PJC, PLC, POT, PTS, S99, SD3, STP, SVC, TKU, TPH,
 BBS, BMI, CMC, DAE, DHI, EBS, HAI, ICF, NLC, NTP, PAN, PJC, S99, SAP, SD3, SD7, SDA, SDT, SGD, SIC, SVC, TKU,
 BMI, BTS, BVS, CMC, GHA, HAI, HPC, ICF, NBC, PAN, PJC, PLC, POT, PTS, S91, S99, SAP, SD6, SD7, SD9, SDA, SDT,
 ACB, BMI, BVS, CJC, CMC, EBS, NBC, NTP, PJC, PPG, S64, S99, SAP, SCC, SD7, SD9, SDA, SJE, SNG, SSI, STP, TLT,
 ACB, BCC, BMI, BTS, BVS, CMC, CTB, CTN, DHI, HJS, HNMM, HPC, ICF, MPC, NBC, NTP, PAN, PLC, POT, S64, S91, S99,
 ACB, BCC, BMI, BTS, BVS, CMC, DHI, EBS, HAI, HNMM, HPC, HTP, ICF, MPC, NBC, NST, NTP, PAN, PLC, POT, PTC, PTS,
 BCC, BMI, BTS, BVS, CTN, DHI, GHA, HAI, HJS, HNMM, HPC, ICF, NBC, PAN, PSC, PTC, S91, SD3, SD7, SDA, SDT, SJE,
 BCC, BMI, BTS, BVS, CTN, DHI, GHA, HJS, HNMM, HPC, LTC, MCO, NBC, NTP, PAN, PPG, PSC, PTC, PTS, SD3, SD7, SDT,
 ACB, BBS, BCC, BMI, BTS, BVS, CTB, CTN, EBS, GHA, HAI, HJS, HNMM, HPC, NBC, NLC, NST, NTP, PAN, PJC, PLC, PPG,
 BBS, BCC, BTS, BVS, CTN, DHI, EBS, GHA, HNMM, HPC, HPS, NBC, NTP, PAN, PJC, PLC, PTS, SD9, SDA, SJE, SSI, STC,
 ACB, BBS, BCC, BTS, BVS, EBS, GHA, HJS, HPC, HPS, HTP, NBC, NLC, NPS, NTP, PAN, PJC, PLC, POT, PPG, PTS, SD3,
 ACB, BBS, BCC, BMI, BTS, BVS, CIC, CTN, DHI, GHA, HJS, HPC, HTP, NBC, NLC, NTP, PAN, PLC, POT, PTS, SCC, SD9,
 ACB, BBS, BCC, BMI, BTS, CJC, CMC, CTB, CTN, DAC, DHI, EBS, GHA, HJS, HPS, HTP, NBC, NLC, NTP, PAN, PLC, POT,
 ACB, BBS, CMC, DAC, DHI, EBS, HJS, HPS, HTP, LTC, NBC, NLC, PPG, PTC, PTS, S91, SCC, SD6, SDA, SDC, SGD, SIC,
 ACB, BVS, CMC, CTB, DAC, DAE, DHI, EBS, HTP, LTC, MPC, NPS, PTC, SAP, SDA, SDC, SDT, SGD, SSI, STP, TKU, TLT,
 ACB, BBS, BCC, BTS, BVS, CTB, CTN, DAC, DAE, DHI, EBS, GHA, HNMM, HPC, HPS, HTP, ICF, MPC, NPS, NST, NTP, PAN,
 ACB, BBS, BCC, BMI, BVS, CMC, CTB, DAE, DHI, EBS, HJS, HLY, HPS, HTP, ICF, NBC, NPS, PJC, PPG, S91, SAP, SD6,
 ACB, BTS, BVS, CMC, DAC, DAE, DHI, EBS, GHA, HLY, HPC, HPS, HTP, ICF, NST, NTP, PAN, PJC, POT, S91, S99, SAP,
 ACB, BTS, BVS, CIC, CID, CMC, DAE, DHI, HAI, HLY, HPC, HPS, ICF, NST, POT, PPG, S64, S99, SAP, SD5, SD6, SD9,
 CIC, CID, CMC, CTB, CTN, DAE, HAI, HJS, HSC, HTP, ICF, LTC, NPS, NST, PLC, POT, PTC, S64, SD7, SD9, SDC, SIC,
 BMI, CIC, CID, CTB, GHA, HLY, PAN, PLC, PPG, PTC, S91, SD7, SD9, SDA, SIC, SJE, SSI, STP, SVC, TBC, TKU, TLT,
 ACB, BBS, BCC, BMI, BTS, BVS, CIC, DAE, DHI, EBS, GHA, HJS, HPC, HSC, MPC, NST, NTP, PAN, PJC, PLC, POT, PPG,

Hình 2. Minh hoạ CSDL tác vụ chứng khoán tăng giá

Cơ sở dữ liệu định lượng được lưu trong bảng Excel như Hình 1 ở trên được tổ chức thành 2 CSDL tác vụ chứng khoán tương ứng là CSDL tác vụ tăng giá và CSDL tác vụ giảm giá. Trong mỗi CSDL tác vụ đó, thông tin về các mã cổ phiếu trong mỗi phiên giao dịch được tổ chức trên cùng một hàng gồm tất cả các mã cổ phiếu của các công ty có giá cổ phiếu tăng (hoặc giảm) trong phiên giao dịch đó. Các mã cổ phiếu được cách nhau bởi dấu phẩy. Khi đó thứ tự của các hàng trong CSDL tác vụ sẽ ngầm định thứ tự (hay mã) của các phiên giao dịch. Hình 2 minh hoạ biểu diễn một phần của CSDL tác vụ chứng khoán về tăng (hoặc

giảm) giá như vậy. Các CSDL tác vụ đó có thể được lưu dưới dạng tệp văn bản và phần mềm CBA hoàn toàn có thể đọc được dữ liệu ở dạng như vậy.

3.2. Phát hiện luật kết hợp chứng khoán từ cơ sở dữ liệu tác vụ

a. Phát hiện luật kết hợp từ cơ sở dữ liệu tác vụ chứng khoán tăng giá

Như đã biết số lượng các luật kết hợp tìm được là phụ thuộc vào độ hỗ trợ cực tiểu và độ tin cậy cực tiểu. Mỗi tổ hợp 2 giá trị này khác nhau sẽ xác định được số lượng các luật kết hợp khác nhau.

Nếu chọn độ hỗ trợ cực tiểu là 35% và độ tin cậy cực tiểu là 70%, số lượng luật kết hợp được phát hiện từ CSDL tác vụ chứng khoán tăng giá thông qua phần mềm CBA là 211 luật. Phân tích 211 luật này cho thấy:

- Có 16 luật kết hợp có Hastc Index nằm ở phần hệ quả, trong đó 5 luật có độ tin cậy cao nhất sắp theo thứ tự giảm dần là:

Rule 4: $HPC = Y, ACB = Y,$

$\rightarrow Hastc Index = Y$ (39.572% 90.54% 74 67 35.829%).

Rule 37: $TLC = Y, HPC = Y,$

$\rightarrow Hastc Index = Y$ (41.711% 85.90% 78 67 35.829%).

Rule 35: $PC = Y \rightarrow Hastc Index = Y$ (53.476% 81.00% 100 81 43.316%).

Rule 33: $SSI = Y \rightarrow Hastc Index = Y$ (51.872% 80.41% 97 78 41.711%).

Rule 26: $BVS = Y \rightarrow Hastc Index = Y$ (47.059% 79.55% 88 70 37.433%).

Luật kết hợp Rule 4 nói rằng 35,829% trong tổng số các phiên giao dịch có các mã cổ phiếu HPC, ACB tăng giá và *HastcIndex* tăng điểm, và với độ tin cậy là 90,54% khi HPC và ACB cùng tăng giá thì Hastc Index tăng điểm. Như vậy 5 luật nêu trên cho biết Hastc Index sẽ tăng điểm khi những nhóm mã chứng khoán sau đây: (HPC,ACB); (TLC, HPC); PC; SSI; BVS tăng giá với độ tin cậy không thấp hơn 79,55%, đó là độ tin cậy của luật Rule 26. Ngoài 5 luật trên, các luật còn lại trong 16 luật có Hastc Index ở phần hệ quả đều có phần tiền đề chỉ gồm có 01 mã cổ phiếu. Theo độ giảm dần của độ tin cậy của luật kết hợp, các phần tiền đề tương ứng với các luật này là các mã cổ phiếu sau: ACB; PAN; TLC; BMI; SVC; VFR; NTP; POT; NLC; HNM; ICF, trong đó luật có độ tin cậy thấp nhất trong 16 luật trên là:

Rule 40: $ICF = Y \rightarrow HastcIndex = Y$ (51.337% 70.83% 96 68 36.364%).

- Phân tích 211 luật kết hợp phát hiện được cho thấy cũng có 16 luật kết hợp có Hastc Index nằm ở phần tiền đề, trong đó 5 luật có độ tin cậy cao nhất sắp theo thứ tự giảm dần là:

Rule 5: $HastcIndex = Y, ACB = Y,$

$\rightarrow HPC = Y$ (42.246% 84.81% 79 67 35.829%).

Rule 3: $HPC = Y, HastcIndex = Y,$

$\rightarrow ACB = Y$ (43.316% 82.72% 81 67 35.829%).

Rule 39: $HPC = Y, HastcIndex = Y,$

$\rightarrow TLC = Y$ (43.316% 82.72% 81 67 35.829%).

Rule 36: $HastcIndex = Y \rightarrow HPC = Y$ (53.476% 81.00% 100 81 43.316%).

Rule 1: $HastcIndex = Y \rightarrow ACB = Y$ (53.476% 79.00% 100 79 42.246%).

Trong đó có 3 luật đầu có nghĩa rằng khi *Hastc Index* tăng đi kèm với một số mã cổ phiếu khác ở phần tiền đề của luật tăng giá có tác động đến sự tăng giá của mã cổ phiếu nằm ở phần hệ quả của luật. Hai luật cuối cùng trong 5 luật tiêu biểu ở trên đều cho thấy chỉ có *Hastc* nằm ở phần tiền đề, điều đó có nghĩa là khi *Hastc* tăng đi kèm thì các mã cổ phiếu trong phần hệ quả của luật sẽ tăng giá. Thật ra 11 luật kết hợp có *Hastc Index* ở phần tiền đề còn lại đều có chung tính chất là phần tiền đề chỉ có mỗi *Hastc Index*, không còn mã chứng khoán nào khác. Và như vậy có 13 luật chỉ có *Hastc Index* ở phần tiền đề và phần hệ quả là các nhóm mã cổ phiếu sau HPC; ACB; SSI; TLC; SD9; NTP; POT; TBC; VMC; SVC; NLC; BVS; HNM. Nói cách khác, khi *Hastc Index* tăng đi kèm thì các nhóm mã cổ phiếu này sẽ tăng giá với độ tin cậy không thấp hơn 70%, đó là độ tin cậy của luật Rule 45, là luật có độ tin cậy thấp nhất trong 16 luật có *Hastc Index* ở phần tiền đề.

Rule 45: $HastcIndex = Y \rightarrow HNM = Y$ (53.476% 70.00% 100 70 37.433%).

- 179 luật còn lại trong tổng số 211 luật kết hợp được phát hiện từ CSDL tác vụ chứng khoán tăng giá, phản ánh mối hệ giữa các nhóm công ty với nhau có thể là cùng thuộc hoặc không một nhóm ngành kinh tế. 8 luật có độ tin cậy cực tiểu cao nhất trong các luật đó là:

Rule 74: $PTC = Y, TBC = Y,$
 $\rightarrow TLC = Y$ (40.107% 89.33% 75 67 35.829%).

Rule 80: $HNM = Y, TBC = Y,$
 $\rightarrow TLC = Y$ (39.572% 89.19% 74 66 35.294%).

Rule 57: $TLC = Y, POT = Y,$
 $\rightarrow TBC = Y$ (40.642% 88.16% 76 67 35.829%).

Rule 72: $TLC = Y, PTC = Y,$
 $\rightarrow TBC = Y$ (41.711% 85.90% 78 67 35.829%).

Rule 59: $POT = Y, TBC = Y,$
 $\rightarrow TLC = Y$ (42.781% 83.75% 80 67 35.829%).

Rule 208: $NM = Y \rightarrow TLC = Y$ (52.406% 82.65% 98 81 43.316%).

Rule 204: $PTC = Y \rightarrow TLC = Y$ (50.802% 82.11% 95 78 41.711%).

Rule 58: $TLC = Y, TBC = Y,$
 $\rightarrow POT = Y$ (43.850% 81.71% 82 67 35.829%).

Trong đó, 2 luật có độ cực tiểu thấp nhất là:

Rule 167: $HPC = Y \rightarrow ICF = Y$ (53.476% 70.00% 100 70 37.433%).

Rule 176: $HPC = Y \rightarrow MPC = Y$ (53.476% 70.00% 100 70 37.433%).

b. Phát hiện luật kết hợp từ cơ sở dữ liệu tác vụ chứng khoán giảm giá

Một cách tương tự, từ CSDL tác vụ chứng khoán giảm giá nếu chọn độ hỗ trợ cực tiểu và độ tin cậy cực tiểu tương ứng là 30% và 60%, qua phần mềm CBA sẽ phát hiện được 88 luật kết hợp, trong đó:

- Có 10 luật kết hợp có *Hastc Index* nằm ở phần hệ quả, và phần tiền đề của các luật này đều chỉ có một mã chứng khoán. Luật có độ tin cậy cao nhất và thấp nhất trong 10 luật đó là các luật:

Rule* 17: $HPC = Y \rightarrow HastcIndex = Y$ (42.246% 82.28% 79 65 34.759%),

Rule* 25: $SDA = Y \rightarrow HastcIndex = Y$ (49.198% 61.96% 92 57 30.481%).

Phần tiền đề của các luật còn lại là các mã chứng khoán sau: SSI; ACB; PAN; BVS; SD9; NTP; BMI; VC2; SDA. Nói cách khác, trên 30% tổng số các phiên giao dịch được chọn thì

một trong các mã cổ phiếu HPC; SSI; ACB; PAN; BVS; SD9; NTP; BMI; VC2, SDA cùng Haste Index giảm giá và giảm điểm, và với độ tin cậy là trên 60%, khi các mã cổ phiếu này giảm giá, Haste sẽ giảm điểm.

- Cũng có 10 luật kết hợp có và chỉ có Haste Index ở phần tiền đề, phần hệ quả cũng chỉ có một mã cổ phiếu, đó là các mã HPC; SSI; ACB; BVS; BMI; SD9; PAN; VC2; NTP và SDA được sắp xếp theo thứ tự giảm dần về độ tin cậy của luật kết hợp có mã cổ phiếu đó ở phần hệ quả. Các luật có độ tin cậy cao nhất và thấp nhất trong 10 luật này là:

Rule* 18: $HasteIndex = Y \rightarrow HPC = Y$ (46.524% 74.71% 87 65 34.759%),

Rule* 26: $HasteIndex = Y \rightarrow SDA = Y$ (46.524% 65.52% 87 57 30.481%).

- Còn lại 68 luật kết hợp mà phần tiền đề và phần hệ quả không chứa Haste Index. 5 luật có độ tin cậy cực tiểu cao nhất và luật có độ tin cậy cực tiểu thấp nhất trong 68 luật này là:

Rule* 46: $S91 = Y \rightarrow SJE = Y$ (40.107% 78.67% 75 59 31.551%).

Rule* 30: $BCC = Y \rightarrow BTS = Y$ (43.316% 77.78% 81 63 33.690%).

Rule* 79: $SD6 = Y \rightarrow SD7 = Y$ (44.385% 77.11% 83 64 34.225%).

Rule* 51: $HPC = Y \rightarrow SSI = Y$ (42.246% 73.42% 79 58 31.016%).

Rule* 77: $SDT = Y \rightarrow SD7 = Y$ (45.455% 72.94% 85 62 33.155%).

Rule* 71: $S99 = Y \rightarrow SD7 = Y$ (53.476% 61.00% 100 61 32.620%).

Ý nghĩa của các luật này cũng được phân tích tương tự như trên.

c. Kiểm định thực tiễn

Để kiểm nghiệm các luật kết hợp phát hiện được, bài báo đã tiến hành lựa chọn dữ liệu kiểm định như sau: chọn 5 phiên giao dịch vào mỗi ngày thứ ba trong tuần của 5 tuần liên tiếp kể từ ngày 11/9/2007. Như vậy 5 ngày được tiến hành kiểm định là: 11/09/2007, 18/09/2007, 25/09/2007, 2/10/2007, 9/10/2007. Trong 5 ngày đó thì ngẫu nhiên chỉ số Haste Index đều tăng điểm so với ngày trước đó.

Theo phân tích ở mục 3.2.a, khi Haste Index tăng điểm thì nói chung nhiều mã trong 13 mã cổ phiếu HPC; ACB; SSI; TLC; SD9; NTP; POT; TBC; VMC; SVC; NLC; BVS; HNM sẽ tăng giá. Kiểm định phân tích này với thực tế ta thấy: ngày 11/9/2007 có các cổ phiếu: HPC; TLC; NTP; POT; VMC; SVC; BVS; HNM tăng giá.; ngày 18/9/2007 các cổ phiếu: HPC; ACB; TLC; SD9; NTP; TBC; VMC; SVC tăng giá còn SSI; POT; NLC; BVS không thay đổi giá; ngày 25/9/2007 và ngày 2/10/2007 tất cả 13 mã cổ phiếu nói trên đều tăng giá và ngày 9/10/2007 có các cổ phiếu: ACB; SSI; NTP; VMC; SVC; BVS tăng giá. Như vậy có thể nói rằng trong 5 phiên giao dịch được lựa chọn ngẫu nhiên này, việc tăng điểm của Haste Index cơ bản là tiền đề để 13 mã cổ phiếu được nêu trên tăng giá.

Ngược lại cũng theo phân tích trên, việc tăng giá của một số nhóm mã cổ phiếu trong các nhóm mã: (HPC, ACB); (TLC, HPC); PC; SSI; BVS; ACB; PAN; TLC; BMI; SVC; VER; NTP; POT; NLC; HNM; ICF sẽ là tiền đề để Haste Index tăng điểm. Thực tế kết quả của 5 phiên giao dịch được lựa chọn ngẫu nhiên như trên cũng cho thấy nhận định này là khá chính xác.

Mặc dù ở 5 phiên giao dịch trên, Haste Index tăng điểm và nhiều mã cổ phiếu tăng giá, nhưng cũng có nhiều mã cổ phiếu giảm giá hoặc đứng giá. Ngày 11/9/2007 có 9 mã cổ phiếu giữ nguyên giá và 20 mã cổ phiếu giảm giá. Ngày 18/9/2007 có 8 mã cổ phiếu giữ nguyên giá và 24 mã cổ phiếu giảm giá. Ngày 25/9/2007 có 6 mã cổ phiếu giảm giá. Ngày 2/10/2007

có 1 mã cổ phiếu giữ nguyên giá và 8 mã cổ phiếu giảm giá. Đặc biệt ngày 9/10/2007 có 3 mã cổ phiếu giữ nguyên giá và có tới 53 mã cổ phiếu giảm giá.

Một cách tương tự khi so sánh mối quan hệ về việc tăng giá của các nhóm mã cổ phiếu với nhau với thực tiễn của 5 phiên giao dịch được lựa chọn cũng cho ta kết luận tương tự như đối với trường hợp của Hasc Index.

Trong số những ngày có Hasc Index giảm điểm của tháng 10, việc chọn ngẫu nhiên 2 phiên giao dịch để kiểm định đã rơi vào các ngày 5/10/2007 và 18/10/2007. Thực tế cho thấy ngày 5/10/2007 các mã cổ phiếu: SSI; ACB; PAN; BVS; NTP; BMI giảm giá và ngày 18/10/2007 các mã cổ phiếu: SSI; CB; PAN; BVS; SD9; NTP; BMI; VC2; SDA giảm giá.

Theo Mục 3.2.b, khi Hasc Index giảm điểm thì nói chung nhiều trong số 13 mã cổ phiếu là: HPC; ACB; SSI; TLC; SD9; NTP; POT; TBC; VMC; SVC; NLC; BVS; HNM giảm giá và ngược lại khi có một số mã cổ phiếu trong các mã: ACB; PAN; TLC; BMI; SVC; VFR; NTP; POT; NLC; HNM; ICF giảm giá thì Hasc giảm điểm. So sánh với thực tế xảy ra của thị trường chứng khoán ở 2 ngày 5/10/2008 và 18/10/2008 ta cũng thấy rằng các luật kết hợp liên quan giữa việc giảm điểm của Hasc Index và giảm giá của một số mã cổ phiếu đã phản ánh khá sát với diễn biến của thị trường.

Tương tự ta cũng có kết luận như vậy đối với các luật kết hợp biểu thị mối quan hệ về sự giảm giá giữa các nhóm mã cổ phiếu khác nhau.

d. Một số lưu ý khi phân tích các luật được phát hiện

- Trong cả 2 trường hợp CSDL chứng khoán tăng giá và giảm giá, đều phát hiện được những luật có đặc điểm là: nhóm mã cổ phiếu này là tiền đề của nhóm mã cổ phiếu kia trong một luật kết hợp nào đó thì lại trở thành hệ quả của nhóm mã cổ phiếu kia trong một luật kết hợp khác. Chẳng hạn như các mã cổ phiếu trong nhóm các luật (Rule 33, Rule 34); (Rule 81, Rule 82); (Rule*17, Rule* 18); hay (Rule*19, Rule* 20),..., là như vậy trong đó:

Rule 34: $HascIndex = Y \rightarrow SSI = Y$ (53.476% 78.00% 100 78 41.711%).

Rule 81: $TLC = Y \rightarrow TBC = Y$ (55.080% 79.61% 103 82 43.850%).

Rule 82: $TBC = Y \rightarrow TLC = Y$ (55.615% 78.85% 104 82 43.850%).

Rule* 17: $HPC = Y \rightarrow HascIndex = Y$ (42.246% 82.28% 79 65 34.759%).

Rule* 19: $SSI = Y \rightarrow HascIndex = Y$ (44.385% 78.31% 83 65 34.759%).

Rule* 18: $HascIndex = Y \rightarrow HPC = Y$ (46.524% 74.71% 87 65 34.759%).

Rule* 20: $HascIndex = Y \rightarrow SSI = Y$ (46.524% 74.71% 87 65 34.759%).

Còn Rule 33 như đã được giới thiệu ở phần trên. Khi đó có thể kết luận rằng khi mã cổ phiếu của nhóm này tăng giá (trường hợp luật được phát hiện từ CSDL tác vụ chứng khoán tăng) hoặc giảm giá (trường hợp CSDL chứng khoán giảm giá) thì nhóm kia cũng vậy và ngược lại.

Đặc biệt cũng tồn tại nhóm mã cổ phiếu là tiền đề cho nhóm mã cổ phiếu khác cả trong 2 trường hợp của CSDL tác vụ chứng khoán và ngược lại nó lại là hệ quả của nhóm mã cổ phiếu kia trong cả hai trường hợp CSDL tác vụ như vậy thì có thể kết luận rằng các nhóm mã cổ phiếu này có quan hệ rất mật thiết với nhau, luôn cùng tăng hoặc cùng giảm giá. Chẳng hạn nhóm mã cổ phiếu trong các luật (Rule 33, Rule 34, Rule*19, Rule* 20) có đặc điểm như vậy.

Trường hợp nhóm mã cổ phiếu này là tiền đề của nhóm mã cổ phiếu kia nhưng ngược lại chỉ có một phần của nhóm mã cổ phiếu kia là tiền đề cho một phần của nhóm đó thì cần kết

hợp thêm một số luật kết hợp khác nữa để có thể đưa ra kết quả tương tự. Kỹ thuật kết hợp nhiều luật kết hợp để sinh ra luật kết hợp mới với độ hỗ trợ và độ tin cậy của luật được ước lượng theo ngưỡng đã được tác giả nghiên cứu và trong bài báo này chưa ứng dụng kỹ thuật kết hợp đó.

- Phát hiện được những luật có một nhóm mã cổ phiếu nào đó là tiền đề của một nhóm mã cổ phiếu khác cho cả hai trường hợp tăng giá và giảm giá, chẳng hạn như mã cổ phiếu trong các luật Rule 33 và Rule*19,..., khi đó có thể kết luận rằng nhóm cổ phiếu này tăng hoặc giảm giá sẽ tác động một cách tương ứng đến việc tăng hoặc giảm giá của nhóm mã cổ phiếu kia.

- Việc phân tích các luật được phát hiện xét theo góc nhìn kinh tế là nội dung cốt yếu nhất khi phân tích các luật kết hợp chứng khoán.

Từ việc xác định được nhóm mã cổ phiếu có tác động đến việc tăng (hoặc giảm) điểm của Hasc Index, thì cần giải mã cho nhóm cổ phiếu đó thuộc ngành kinh tế nào? tổng giá trị của công ty tính theo chứng khoán là bao nhiêu? kinh doanh lỗ lãi những năm qua ra sao? hệ số P/E thế nào?,... Những phân tích ấy sẽ giúp nhà hoạch định chính sách và đầu tư có cách nhìn sát thực và lý giải nguyên nhân tăng hoặc giảm điểm của Hasc Index. Chẳng hạn như đã trình bày ở trên Hasc Index chịu tác động mạnh của các mã cổ phiếu như SSI, ACB; PAN; TLC; BMI; SVC; VFR; NTP; POT; NLC; HNM; ICF, đó là mã cổ phiếu tương ứng của Công ty cổ phần chứng khoán Sài Gòn (SSI), Ngân hàng Á châu (ACB),..., Đó đều là những đơn vị phát hành số lượng lớn cổ phiếu, cổ phiếu có giá trị cao và được giao dịch nhiều. Đặc biệt là mã cổ phiếu SSI. Hasc Index và SSI có quan hệ mật thiết với nhau như nhận xét ở trên, khi SSI tăng hoặc giảm giá thì Hasc Index cũng tăng hoặc giảm điểm và ngược lại và như chúng ta đã biết khi SSI niêm yết trên sàn Hà Nội, khối lượng giao dịch của SSI thường chiếm từ 60%-70% khối lượng giao dịch của thị trường.

Từ việc giải mã các nhóm mã cổ phiếu theo cách nhìn kinh tế, ta sẽ có thêm căn cứ để lý giải nguyên nhân tại sao các nhóm mã cổ phiếu này cùng tăng hoặc cùng giảm giá, phải chăng các mã cổ phiếu trong các nhóm đó có quan hệ với nhau về tổ chức theo một nghĩa nào đó? Đã có những chính sách khuyến khích thúc đẩy đầu tư, phát triển những nhóm ngành của các công ty có mã cổ phiếu đó? hay sản phẩm của các công ty này có liên quan đến nhau và cầu của chúng ngày càng tăng,... Chẳng hạn từ các luật kết hợp chứng khoán giảm giá:

Rule* 56: $PAN = Y \rightarrow SSI = Y$ (43.850% 70.73% 82 58 31.016%),

Rule* 57: $SSI = Y \rightarrow PAN = Y$ (44.385% 69.88% 83 58 31.016%).

Ta thấy SSI và PAN có quan hệ khá mật thiết với nhau, nhưng ít người biết là Chủ tịch Hội đồng quản trị SSI đã khởi nghiệp từ PAN và đến nay PAN vẫn được coi là công ty "chị em" của SSI, nếu không biết như vậy có lẽ sẽ rất khó giải thích được về mức tăng trưởng "chóng mặt" giá trị cổ phiếu của một công ty vệ sinh bình thường như PAN thời gian qua.

4. KẾT LUẬN

Luật kết hợp đã được sử dụng khá thành công trong khai phá dữ liệu, nhưng trong lĩnh vực chứng khoán thì có thể nói là hầu như chưa có. Bài báo này cho thấy phát hiện luật kết hợp thực sự là một công cụ phân tích kỹ thuật mới, có ích không chỉ cho các nhà đầu tư mà còn cho các nhà quản lý và hoạch định chính sách về phát triển thị trường chứng khoán.

Do thị trường chứng khoán tập trung ở Việt Nam mới chỉ phát triển hơn 2 năm trở lại

đây và còn có ít cổ phiếu của các công ty được niêm yết ở các trung tâm giao dịch chứng khoán do vậy số liệu thu thập được chưa đủ lớn vì thế các luật kết hợp phát hiện được chưa thể phản ánh được đầy đủ các mối quan hệ giữa các mã cổ phiếu với nhau và giữa các mã cổ phiếu với Hasc Index. Tuy nhiên cùng với sự phát triển của thị trường chứng khoán Việt Nam, kho dữ liệu chứng khoán sẽ lớn lên nhanh chóng và khi đó việc ứng dụng kỹ thuật phát hiện luật kết hợp để phân tích thị trường chứng khoán hứa hẹn sẽ trở thành giải pháp rất hữu ích do kỹ thuật này đặc biệt có ưu thế khi kho dữ liệu lớn hoặc rất lớn.

Bài báo này chỉ được xem là nghiên cứu ban đầu nhằm trả lời những câu hỏi được đề xuất. Còn rất nhiều vấn đề khác đang đặt ra cần được quan tâm nghiên cứu, chẳng hạn quan hệ về mức độ tăng (hoặc giảm) giá của các mã cổ phiếu ở phần tiền đề và phần hệ quả của các luật kết hợp chứng khoán Việt nam thế nào ngoài việc tăng (hoặc giảm) giá của nhóm mã cổ phiếu ở phần tiền đề tác động đến việc tăng (hoặc giảm) điểm của Hasc Index thì số lượng cổ phiếu, giá trị cổ phiếu được bán ra trong ngày của các mã cổ phiếu ở phần tiền đề chắc chắn cũng tác động không nhỏ đến mức độ tăng (hoặc giảm) điểm của Hasc Index, vấn đề là làm cách nào phát hiện được quan hệ đó giữa chúng ?, ...Bài báo này chưa giải quyết những vấn đề như vậy. Đối với những vấn đề nêu trên, kỹ thuật phát hiện luật kết hợp định lượng hoặc luật kết hợp mờ từ CSDL định lượng có nhiều hứa hẹn sẽ là giải pháp tốt và đó cũng là hướng nghiên cứu trong tương lai gần của tác giả bài báo.

Bài báo cũng chưa tìm được các luật kết hợp từ CSDL tác vụ ở đó phần tiền đề, cũng như phần hệ quả gồm hỗn hợp các mã cổ phiếu, trong đó có mã tăng giá, có mã giảm giá, có mã đứng giá. Thật ra vấn đề phát hiện luật kết hợp hỗn hợp như vậy từ các CSDL tác vụ hiện vẫn đang là vấn đề mở, tuy rằng nó được khá nhiều người quan tâm nghiên cứu nhưng kết quả còn rất hạn chế, mới dừng ở một số kết quả nghiên cứu ban đầu về bản chất toán học của dạng luật như vậy và chưa có thuật toán hiệu quả để phát hiện được những luật đó.

TÀI LIỆU THAM KHẢO

- [1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen H, Fast discovery of association rules, *Advances in Knowledge discovery and DataMining*, edited by U.M. fayyad, G.Platstsky-Shapiro, P.Smyth, and Uthurusamy, AAAI Press/The MIT Press, 1996 (306-328).
- [2] Boris Kovalerchuk - Evgenii Vityaev, *Data Mining in Finance - Advances in Relational and Hybrid Methods*, Kluwer Academic publisher, 2001.
- [3] A. Gyenesei, "A Fuzzy Approach for Mining Quantitative Association Rules", Turku Centre for Computer Sciences, TUCS Technical Report, No 336, 2000.
- [4] M. Kuod, P. Ada, Mining fuzzy association rules, *SIGMOD Record* **27** (1) (1998).
- [5] Đỗ Văn Thành, "Lý thuyết toán học của phát hiện luật kết hợp", Bài giảng Trường thu hệ mờ và ứng dụng, Viện Toán học, Hà Nội 9/2003.
- [6] Lê Thị Mai Linh, *Phân tích và đầu tư chứng khoán*, Nhà xuất bản Chính trị Quốc gia, 2003.
- [7] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, Efficient mining of association rules using closed itemset latics, *Information Systems* **24** (1) (1999) 20-46.

- [8] M. J. Zaki and M. Ogihara, Theoretical foundation of association rules, 3rd *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, June 1998.
- [9] Ebook - *MetaStock Professional - Equis International* - 2002.
- [10] www.hastc.org.vn.
- [11] www.bsc.com.vn.
- [12] www.vndirect.com.vn.
- [13] www.nus.com.

Nhận bài ngày 11 - 12 - 2007