Association Mining on Stock Index Indicators

Krittithee Utthammajai and Pakorn Leesutthipornchai

Abstract—The stock index indicators, e.g., Relative Strength Index (RSI), Price Rate of Change (PROC), Moving Average Convergence/Divergence (MACD), Stochastic Oscillator (STH), or Chaikin Oscillator (CHK) are often used as the main factor in trading. Stock traders make a decision to trade by using these indicators. This paper applies the association technique to find the hidden relationship among those stock indicators and the volume which affects the market price. The transactional data are captured from Thai stock market in the period from 9 April 2013 to 16 April 2014. The association results show that the set of indicators and volume affects the price change. When the indicators are changed substantially, the price will change significantly.

Index Terms-Association rule, indicator, stock index.

I. INTRODUCTION

Stock index trading is one way for making profits. Professional traders invest their money and gain more money than an interest from bank saving. Traders use previous and current trading data to predict the trends of stock prices in the near future. For example, traders use the ratio of current bid (buying) and offer (selling) volumes and the current price to predict the future price. If the amount of bid volume is high while the offer volume is low and the current price is low. The future price is increased with high probability.

The stock index indicator is the financial formula, e.g., Relative Strength Index (RSI), Price Rate of Change (PROC), Moving Average Convergence/Divergence (MACD), Stochastic Oscillator (STH), or Chaikin Oscillator (CHK). Traders make a technical analysis by using these indicators which the contiguous of indicator values are used to predict the trend of market prices.

This paper considers the relation among the stock index indicators, trading volume, and the market price. The market data are retrieved from the *Stock Exchange of Thailand (SET)* index. We find-tuned the parameters of the data mining technique to find the association among stock indicators, volume, and price.

The remainder of this paper is organized as follows. Section II discusses related works for applying the association rule for stock index. Section III provides stock index indicator formulation. Section IV shows the SET index data set before and after preprocess. Section V discusses the obtained association results. Lastly, Section VI concludes our research work and describes our contributions.

II. RELATED WORKS

Sung Hoon Na and So Young Sohn [1] proposed the forecasting changes in Korea composite stock price index (KOSPI) using association rules. The research applies the association technique to find the relation of price direction (up or down) of stock market indices in Korea, USA, Europe, and Asia. From their association rules [1], the KOSPI index tends to have the same price direction as stock indices in USA and Europe, whereas, the KOSPI tends to have the opposite price direction to stock indices in East Asian countries such as Hong Kong and Japan.

Anantaporn Srisawat [2] proposed an application of association rule mining based on stock market. The paper applies an association technique to discover the relationship of price direction (up and down) among Thai stock indices. The obtained association rules are used to predict the trend of consequence stock index associated to the antecedence indices of the rule.

Sun Jing *et al.* [3] proposed the study of association rule mining on technical action of ball games. The paper applies the association technique (i.e., Apriori algorithm) to find the relationship among technical actions of ball games.

Jiang Hong-bo and Yang De-li [4] proposed the application research on fast discovery of association rules based on air transportation. The research applies the association technique to discover the relationship of the large complex data in air transportation.

Rakesh Agrawal *et al.* [5] proposed mining association rules between sets of items in large databases. The paper introduces the algorithm to generate association rules from buying items in the database.

Previous research works [1]-[5] applied association mining to find the hidden relationship among the given data. The research works [1], [2] provide the relationship among stock indices in many countries, or the relationship among stock indices in Thai stock market. Both of them use association technique to find the unclear relationship from stock data. This paper enhances the benefit of association technique to find the hidden relationship among stock index indicators, the volume, and the market price.

III. STOCK INDEX INDICATORS

Financial instruments called *indicators* are popularly used by trader for predicting the trends of stock prices. The indicator is calculated from the historical data sets. The examples of stock index indicators are Moving Average (MA), Exponential Moving Average (EMA), and etc. In this paper, we provide the formulation of five indicators such as Relative Strength Index (RSI), Price Rate of Change (PROC), Moving Average Convergence Divergence (MACD), Stochastic

Manuscript received June 25, 2014; revised September 1, 2014.

The authors are with the Computer Science Department, Thammasat University, Rangsit Campus, Pathumthani, Thailand (e-mail: krittithee.u@gmail.com; pakornl@cs.tu.ac.th).

Oscillator (STH), and Chaikin Oscillator (CHK). These indicators are described as follows.

A. Relative Strength Index (RSI)

RSI is developed by J. Welles Wilder. RSI is a momentum oscillator that measures the speed and the change of price movements [6]. RSI shows overbought and oversold signals which helps trader make a trading decision. RSI is calculated by the formula as follows.

$$RSI = 100 - \frac{100}{(1 - RS)} \tag{1}$$

where RS is the average gain over the average loss, the average gain is the summation of gains over the previous periods, and the average loss is the summation of loss over the previous periods [6]. RSI oscillates between 0 and 100.

Traders consider two threshold lines. The overbought signal is generated when the RSI value is greater than 70 and oversold signal is generated when the RSI value is less than 30.

B. Price Rate of Change (PROC)

PROC is a pure momentum oscillator that measures the percentage of change in the price. The PROC calculation compares the current price with the price n periods ago [6]. PROC is defined as follows.

$$PROC = \left[\frac{(C - CNP)}{CNP}\right] \times 100$$
(2)

where *C* is the closing price today and CNP is closing price *n* periods ago. PROC value is greater than zero indicates that the buying pressure is increased. PROC value is less than zero indicates that the selling pressure is increased. Traders observe the PROC with the actual price. For the simple example, if the PROC trends to decrease below the zero line and the actual price trends to increase, the selling signal is generated. In the other hand, if the PROC trends to decrease, the buying signal is generated.

C. Moving Average Convergence/Divergence (MACD)

MACD is developed by Gerald Appel in the late seventies [6]. MACD is used to forecast the overbought (or the oversold) trading. MACD fluctuates above and below the zero line as the moving averages convergence and divergence [6]. MACD is calculated by using the following formula.

$$MACD = (N1 - N2) \tag{3}$$

where *N*1 is the exponential moving average of *N*1 periods and *N*2 is the exponential moving average of *N*2 periods. *N*1 must be shorter than *N*2.

D. Stochastic Oscillator (STH)

STH is developed by George C. Lane in the late 1950s. STH is a momentum indicator that shows the location of the close relative to the high-low range over a set number of periods [6]. STH is used to predict the overbought trading or the oversold trading. STH is calculated as follows.

$$STH = \left[\frac{C-L}{H-L}\right] \times 100 \tag{4}$$

where *C* is the closing price. *L* is the lowest price over *n* periods. *H* is the highest price over *n* periods. STH oscillates between 0 and 100.

Traders consider two threshold lines. The overbought signal is generated when the STH value is greater than 80 and oversold signal is generated when the STH value is less than 20.

E. Chaikin Oscillator (CHK)

CHK is developed by Marc Chaikin. CHK measures the momentum of the *Accumulation Distribution Line* (ADL) [6]. CHK oscillates above and below the zero line. If CHK value is positive, the buying pressure is strong. In the other hand, If CHK value is negative, the selling pressure is strong. CHK is calculated as follows.

$$MFM = \left[\frac{(C-L) - (H-C)}{(H-L)}\right]$$
(5)

$$MFV = MFM \times VP \tag{6}$$

ADL = Previous ADL + Current Period's MFV (7)

$$CHK = (N1_EMA \text{ of ADL}) - (N2_EMA \text{ of ADL}) (8)$$

where *C* is the closing price. *L* is the lowest price. *H* is the highest price. *MFM* is the money flow multiplier. VP is the volume for the period. MFV is the money flow volume. ADL is the accumulation distribution line. *N1_EMA* is the short moving average of ADL. *N2_EMA* is the long moving average of ADL.

There are four steps for calculating the CHK indicator. First, the money flow multiplier (MFM) is calculated by using (5). Second, the money flow volume (MFV) is calculated by using (6). Third, the accumulation distribution line (ADL) is calculated by using (7). Fourth, we calculate the difference between two moving averages of the ADL by using (8).

IV. DATA PROCESSING

A. Data Set

The data set was captured from the site that is usually cited by Thai financial institution, i.e., www.efinancethai.com. We captured the data every half an hour in the period of April 9, 2013 to April 16, 2014. Each record of the data set consists of trading date, trading time, opening, high, low, closing price and the trading volume. The currency of the price is Thai Baht. The example of the data set is showed in Table I.

B. Data Preprocessing

The indicator values (e.g., RSI, PROC) are calculated from the original data set by using the formulas that are described in Section II. Next, each indicator is discretized into four ranges. The data preprocessing steps are described as follows.

Date	Time	Open	High	Low	Close	Vol
12/04/2013	09:30:00	196.5	196.5	195	196	97,400
12/04/2013	10:00:00	196	196.5	195	196	245,100
12/04/2013	10:30:00	196.5	196.5	196	196.5	132,600
12/04/2013	11:00:00	196	197.5	196	197	280,900
12/04/2013	11:30:00	197	197	196.5	196.5	984,100
12/04/2013	12:00:00	196.5	197	196.5	197	251,700
12/04/2013	14:00:00	196.5	197	196.5	197	23,200
12/04/2013	14:30:00	196.5	197.5	196.5	197	224,900
12/04/2013	15:00:00	197	197	196	196.5	422,900
12/04/2013	15:30:00	196.5	197	196	196.5	325,000
12/04/2013	16:00:00	197	197.5	196	196.5	266,700
12/04/2013	16:30:00	197.5	197.5	197.5	197.5	427,100

TABLE I: THE EXAMPLE OF THE DATA SET

1) The indicator values (RSI, PROC, MACD, STH, and CHK), the volume, and the percentage of price change (PCH) are calculated from the original data set. The data set with indicators is showed in Table II.

TABLE II: THE DATA SET WITH INDICATORS

Date	Time	RSI	PROC	MACD	STH	СНК	Vol	PCH
12/04/2013	09:30:00	100.00	2.51	1.64	90.00	- 46,121.42	97,400.00	2.08
12/04/2013	10:00:00	100.00	2.10	1.25	63.64	-109,254.56	245,100.00	1.25
12/04/2013	10:30:00	100.00	2.53	1.04	63.64	-116,178.20	132,600.00	1.25
12/04/2013	11:00:00	100.00	1.67	1.14	70.00	- 83,977.53	280,900.00	1.67
12/04/2013	11:30:00	100.00	3.35	1.25	88.89	102,665.81	984,100.00	2.92
12/04/2013	12:00:00	100.00	2.08	1.39	62.50	313,250.15	251,700.00	2.08
12/04/2013	14:00:00	100.00	0	1.46	50.00	343,793.21	23,200.00	2.08
12/04/2013	14:30:00	100.00	1.65	1.29	83.33	203,102.14	224,900.00	2.92
12/04/2013	15:00:00	100.00	0.82	1.21	50.00	- 1,452.34	422,900.00	2.08
12/04/2013	15:30:00	100.00	0.82	1.14	60.00	- 88,370.59	325,000.00	2.50
12/04/2013	16:00:00	100.00	- 0.81	0.93	40.00	-137,264.81	266,700.00	2.08
12/04/2013	16:30:00	100.00	0.82	0.68	75.00	-147,458.32	427,100.00	2.92

2) Since each attribute value in Table II has a large number of distinct values. Therefore this paper reduces a large number of distinct values by using a standard data mining technique (i.e., discretization). Each data record is divided into four ranges (a, b, c, d) as shown in Table III.

-										
aRSI		< 25.00	aPROC	<	-7.00	aMACD	< -4.02		aSTH	< 25.00
bRSI	25.0)0 to 49.99	bPROC	-7.00) to -1.7	0 bMACD	-4.02 to -0.9	0	bSTH	25.00 to 49.99
cRSI	50.0	00 to 75.00	cPROC	-1.7	l to 3.6	0 cMACD	-0.91 to 2.23	3	cSTH	50.00 to 75.00
dRSI	2	> 75.00	dPROC	>	3.60	dMACD	> 2.23		dSTH	> 75.00
aCH	IK	<	101 308 16		aVol	< 1.0	70 750 00		aPCH	< .6.66
aur	II.	N N	91,398.10		avoi	\$1,0	/0,/50.00		aron	<-0.00
bCH	ΙK	-491,398.	16 to 226,2	86.99	bVol	1,070,750.00) to 2,139,70).99	bPCH	-6.66 to -0.98
cCH	łK	226,287.0	00 to 943,97	70.16	cVol	2,139,701.00) to 3,208,650	0.00	cPCH	-0.99 to 4.71
dCH	ΙK	> 9	43,970.16		dVol	> 3,2	08,650.00		dPCH	> 4.71

TABLE III: THE RANGE OF EACH ATTRIBUTE

3) The data set with indicators in Table II is transformed by applying with the data ranges in Table III. Then we obtained the preprocessed data set as shown in Table IV. Each attribute of the data set in Table IV has four distinct values.

From the first record in Table II and Table IV, The date is April 12, 2013 and the time is 9:30. The RSI value from Table

II is 100. The value 100 is in the range dRSI (i.e., more than 75). Therefore the RSI value after preprocessed in Table IV is dRSI.

The PROC value in Table II is 2.51. The value 2.51 is in the range cPROC (i.e., between -1.71 and 3.60). Therefore the PROC value after preprocessed is cPROC and so on in the other indicators.

TABLE IV: THE PREPROCESSED DATA SET

Date	Time	RSI	PROC	MACD	STH	СНК	Vol	PCH
12/04/2013	09:30:00	dRSI	cPROC	cMACD	dSTH	bCHK	aVol	cPCH
12/04/2013	10:00:00	dRSI	cPROC	cMACD	cSTH	bCHK	aVol	cPCH
12/04/2013	10:30:00	dRSI	cPROC	cMACD	cSTH	bCHK	aVol	cPCH
12/04/2013	11:00:00	dRSI	cPROC	cMACD	cSTH	bCHK	aVol	cPCH
12/04/2013	11:30:00	dRSI	cPROC	cMACD	dSTH	bCHK	aVol	cPCH
12/04/2013	12:00:00	dRSI	cPROC	cMACD	cSTH	cCHK	aVol	cPCH
12/04/2013	14:00:00	dRSI	cPROC	cMACD	cSTH	cCHK	aVol	cPCH
12/04/2013	1 <mark>4:30:00</mark>	dRSI	cPROC	cMACD	dSTH	bCHK	aVol	cPCH
12/04/2013	15:00:00	dRSI	cPROC	cMACD	cSTH	bCHK	aVol	cPCH
12/04/2013	15:30:00	dRSI	cPROC	cMACD	cSTH	bCHK	aVol	cPCH
12/04/2013	16:00:00	dRSI	cPROC	cMACD	bSTH	bCHK	aVol	cPCH
12/04/2013	16:30:00	dRSI	cPROC	cMACD	dSTH	bCHK	aVol	cPCH

V. STOCK INDEX RELATIONS

For our experimental design, we varied the percentage of minimum support from 10% to 80% and captured the number of association rules. The minimum confidence is 100%. From Table V, the obtained association rules are separated into four ranges of the percentage of price change (PCH). When the percentage of minimum support is increased, the number of association rules is reduced.

TABLE V: THE NUMBER OF ASSOCIATION RULES

The Percentage of Price Change (PCH)			Mi	nimun S	Support	(%)						
	10	<mark>2</mark> 0	30	40	50	60	70	80				
< -6.66	409	177	63	41	19	9	5	3				
-6.66 to -0.98	154	97	43	27	14	7	5	-				
-0.99 to 4.71	157	69	35	31	24	16	12	6				
> 4.71	233	97	61	32	20	12	5	3				

From the obtained association rules in Table V, this paper focuses on the minimum support threshold at 70% or above. The top five association rules are showed in Table VI.

From Table VI, In the percentage of price change (PCH) is less than -6.66%, the rule number 3 is {RSI=aRSI, STH=aSTH} \rightarrow PCH=aPCH. The meaning is when the RSI indicator is less than 25 and the STH indicator is less than 25. Consequently, the percentage of price change is oscillated less than -6.66%. The rule number 5 is {RSI=aRSI, PROC=bPROC) \rightarrow PCH=aPCH. The meaning is when the RSI indicator is less than 25 and the PROC indicator is between -7.0 and 1.70. The percentage of price change is oscillated less than -6.66%.

The Percentage of Price Change (PCH)	Minsup %	Association Rules	Support %
	8	$1.{RSI=aRSI} \rightarrow PCH=aPCH$	100
		2.{STH=aSTH} \rightarrow PCH=aPCH	80
<-6.66 %	70	3.{RSI=aRSI STH=aSTH }→ PCH=aPCH	80
		4. {PROC=bPROC} \rightarrow PCH=aPCH	70
		5.{RSI=aRSI PROC=bPROC}→ PCH=aPCH	70
		$1.{Vol=aVol} \rightarrow PCH=bCH$	90
		$2{CHK=bCHK} \rightarrow PCH=bPCH$	86
-6.66% to -0.98%	70	3.{CHK=bCHK Vol=aVol} \rightarrow PCH=bPCH	79
		4.{RSI=aRSI}→PCH=bPCH	79
		5.{RSI=aRSI Vol=aVol}→PCH=bPCH	71
	2	1.{Vol=aVol}→PCH=cPCH	94
		2.{PROC=cPROC}→PCH=cPCH	93
-0.99% to 4.71%	80	3.{PROC=cPROC Vol=aVol}→PCH=cPCH	88
		4.{CHK=bCHK}→PCH=cPCH	84
		5.{MACD=cMACD}→PCH=cPCH	81
		1.{RSI=dRSI}→PCH=dPCH	94
> 4.71%		2.{Vol=aVol}→PCH=dPCH	87
	70	3.{RSI=dRSI Vol=aVol}→PCH=dPCH	82
		4.{STH=dSTH}→PCH=dPCH	80
		5.{RSI=dRSI STH=dSTH}→PCH=dPCH	73

TABLE VI: THE ASSOCIATION RULES

In the percentage of price change is between -6.66% and -0.98%, the rule number 3 is {CHK=bCHK, $Vol=aVol\} \rightarrow PCH=bPCH$. The meaning is when the CHK indicator is between -491,398.16 and 226,286.99 and the trading volume is less than 1,070,750. The percentage of price change is oscillated between -6.66% and -0.98%. The rule number 5 is {RSI=aRSI, Vol=aVol} \rightarrow PCH=bPCH. It means when the RSI indicator is less than 25 and the volume is less than 1,070,750. The percentage of price change is oscillated between -6.66% and -0.98%.

In the percentage of price change is between -0.99% and 4.71%, The rule number 2 is {PROC=cPROC} \rightarrow PCH=cPCH. The meaning is when the PROC indicator is between -1.71 and 3.60, the percentage of price change is oscillated between -0.99% and 4.71%. The rule number 5 is {MACD=cMACD} \rightarrow PCH=cPCH. It means when the MACD indicator is between -0.91 and 2.23, the percentage of price change is oscillated between -0.99% and 4.71%.

In the percentage of price change is more than 4.71%, the rule number 3 is {RSI=dRSI, Vol=aVol} \rightarrow PCH=dPCH. The meaning is when the RSI indicator is more than 75 and the stock volume is less than 1,070,750. Consequently, the percentage of price change is oscillated more than 4.71%. The rule number 5 is {RSI=dRSI STH=dSTH} \rightarrow PCH=dPCH. The rule number 5 consists of two antecedence indicators. The meaning is when the RSI indicator is more than 75 and the STH indicator is more than 75, the percentage of price change is oscillated more than 4.71%.

VI. CONCLUSION

This paper applies the association technique to find the hidden relations among stock index indicators and the market volume that affect the percentage of price change. The transactional data are captured from Thai stock market in the period from 9 April 2013 to 16 April 2014. This paper varies the percentage of minimum support from 10% to 80%. The minimum confidence is 100%. The obtained association rules with at least 70% minimum support are considered and described. The association results show that the antecedence indicators are related to the consequence price change. The set of indicators are changed substantially, the price will change significantly.

REFERENCES

- S. H. Na and S. Y. Sohn, "Forecasting changes in Korea composite stock price index (KOSPI) using association rules," *Journal of Expert Systems with Applications*, vol. 38, pp. 9046–9049.
- [2] A. Srisawat, "An application of association rule mining based on stock market," in *Proc. 3rd Inter. Conf. on Data Mining and Intelligent Information Technology Applications*, Oct. 2011, pp. 259-262.
- [3] J. Sun et al., "Study of association rule mining on technical action of ball games," in Proc. Inter. Conf. on Measuring Technology and Mechatronics Automation, 2010, pp. 539-542.
- [4] H. B. Jiang and D. L. Yang, "Application research on fast discovery of association rules based on air transportation," in *Proc. Inter. Conf. on Service Systems and Service Management*, Jun. 2007, pp. 1-6.
- [5] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large data bases," in *Proc. the 1993 ACM SIGMOD Inter. Conf. on Management of Data-SIGMOD*, 1993, pp. 207-216.
- [6] technical indicators and overlays. [Online]. Available: http://www.stockcharts.com/school/doku.php?id=chart_school:techni cal_indicators:chaikin_oscillator



Krittithee Utthammajai was born in Thailand. He received the B.A. degree from Burapha University, Chonburi, Thailand, in 2008. Currently, he studies the M.Sc. degree at the Thammasat University, Rangsit Campus, Pathumthani, Thailand. His research interest is data mining.



Pakorn Leesutthipornchai was born in Thailand in 1983. He received the B.E., M.E., and Ph.D. degrees from King Mongkut's University of Technology Thonburi, Thailand, in 2005, 2007, and 2011, respectively. Since 2011, he has been with the Department of Computer Science, Faculty of Science and Technology, Thammasat University, Thailand, where he is currently a lecturer. His main areas of research interest are data mining, optimization

algorithms and multi-objective optimization.