

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Development of a Classification Rule Mining Framework by Using Temporal Pattern Extraction

Hidenao Abe

*Department of Medical Informatics, Shimane University
Japan*

1. Introduction

In recent years, KDD (Knowledge Discovery in Databases) (Fayyad et al., 1996) has been widely known as a process to extract useful knowledge from databases. In the research field of KDD, 'Temporal (Time-Series) Data Mining' is one of important issues to mine useful knowledge such as patterns, rules, and structured descriptions for a domain expert. However, huge numerical temporal data such as stock market data, medical test data, and sensor data have been only stored to databases. Besides, many temporal mining schemes such as temporal pattern extraction methods and frequent itemset mining methods have been proposed to find out useful knowledge from numerical temporal databases. Although each method can find out partly knowledge of each suggested domains, there is no systematic framework to utilize each given numerical temporal data through whole of the KDD process.

To above problems, we have developed an integrated temporal data mining environment, which can apply numerical temporal data to find out valuable knowledge systematically. The environment consists of temporal pattern extraction, mining, mining result evaluation support system to attempt numerical temporal data from various domains.

In this chapter, we describe a classification rule mining framework by combining temporal pattern extraction and rule mining. This framework has been developed for mining if-then rules consisting of temporal patterns in left hand side of the rules. The right hand side of the rules is indicated to predict both of important events and temporal patterns of important index. In order to show the effectiveness of the framework, we implemented this framework for a medical sequential data of laboratory test results for chronic hepatitis patients and a sequential data consisting of technical indexes for Japanese stocks. By using the implementations and experimental results, we present the following merits achieved by the classification rules with considering temporal patterns of the target attributes:

- Finding different interesting aspects of the decisions/results
- Finding important temporal patterns and attributes for the decision at the same time

In the remaining of this chapter, we describe the related works of this framework in Section 2. In Section 3, we present the framework to mine classification rules that are consisting of temporal patterns and decisions¹. After implementing this framework, an experiment about Japanese stock trading is performed in Section 4. Finally, we summarize the experimental results in Section 6.

¹'decision' means just a cross-sectional decision making, and also means important future situations in this framework.

2. Related work

Many efforts have been done to analyze temporal data at the field of pattern recognitions. Statistical methods such as autoregressive model (Akaike, 1969) and ARMA (Auto Regressive Integrated Moving Average) model have been developed to analyze temporal data, which have linearity, periodicity, and equalized sampling rate. As signal processing methods, Fourier transform, Wavelet (Mallat, 1989), and fractal analysis method have been also developed to analyze such well formed temporal data. These methods based on mathematic models restrict input data, which are well sampled.

However, temporal data include ill-formed data such as clinical test data of chronic disease patients, purchase data of identified customers, and financial data based on social events. To analyze these ill-formed temporal data, we take another temporal data analysis method such as DTW (Dynamic Time Wrapping) (Berndt & Clifford, 1996), temporal clustering with multiscale matching (Hirano & Tsumoto, 2002), and finding Motif based on PAA (Piecewise Approximation Aggregation) (Keogh et al., 2003).

For finding out useful knowledge to decide orders for stock market trading, many studies have done. For example, temporal rule induction methods such as Das's framework (Das et al., 1998) have been developed. Frequent itemset mining methods are also often attempt to the domain (Wong & Fu, 2006). Although they analyze the trend of price movement, many trend analysis indices such as moving average values, Bolinger band signals, MACD signals, RSI and signals based on balance table are often never considered.

In addition, these studies aim not to find out decision support knowledge, which directly indicates orders for stock market trading, but useful patterns to think better decision by a domain expert. Therefore, the decision support of trading order is still costly task even if a domain expert uses some temporal data analysis methods. The reason of this problem is that decision criteria of trading called anomaly are obtained from very complex combination of many kinds of indices related to the market by domain experts.

3. An integrated framework for temporal rule mining by using automatic temporal pattern extraction

Our temporal data mining environment needs temporal data as input. Output rules are if-then rules, which have temporal patterns or/and ordinal clauses, represented in $A = x$, $A \leq y$, and $A > z$. Combinations of extracted patterns and/or ordinal clauses can be obtained as if-then rules by a rule induction algorithm.

To implement the environment, we have analyzed temporal data mining frameworks (Das et al., 1998; Ohsaki et al., 2004). Then, we have identified procedures for pattern extraction as data pre-processing, rule induction as mining, and evaluation of rules with visualized rule as post-processing of mined result. The system provides these procedures as commands for users. At the same time, we have designed a graphical interface, which include data processing, validation for patterns on elemental sequences, and rule visualization as charts.

Our integrated time-series data mining environment combines the following major functional components: time-series data pre-processing, mining, post-processing for mined results, and other database operators to validate data and results of every phase.

With this environment, we aim the following efforts for each agent:

1. Developing and improving time-series data mining procedures for system developers
2. Collaborative data processing and rule induction for data miners

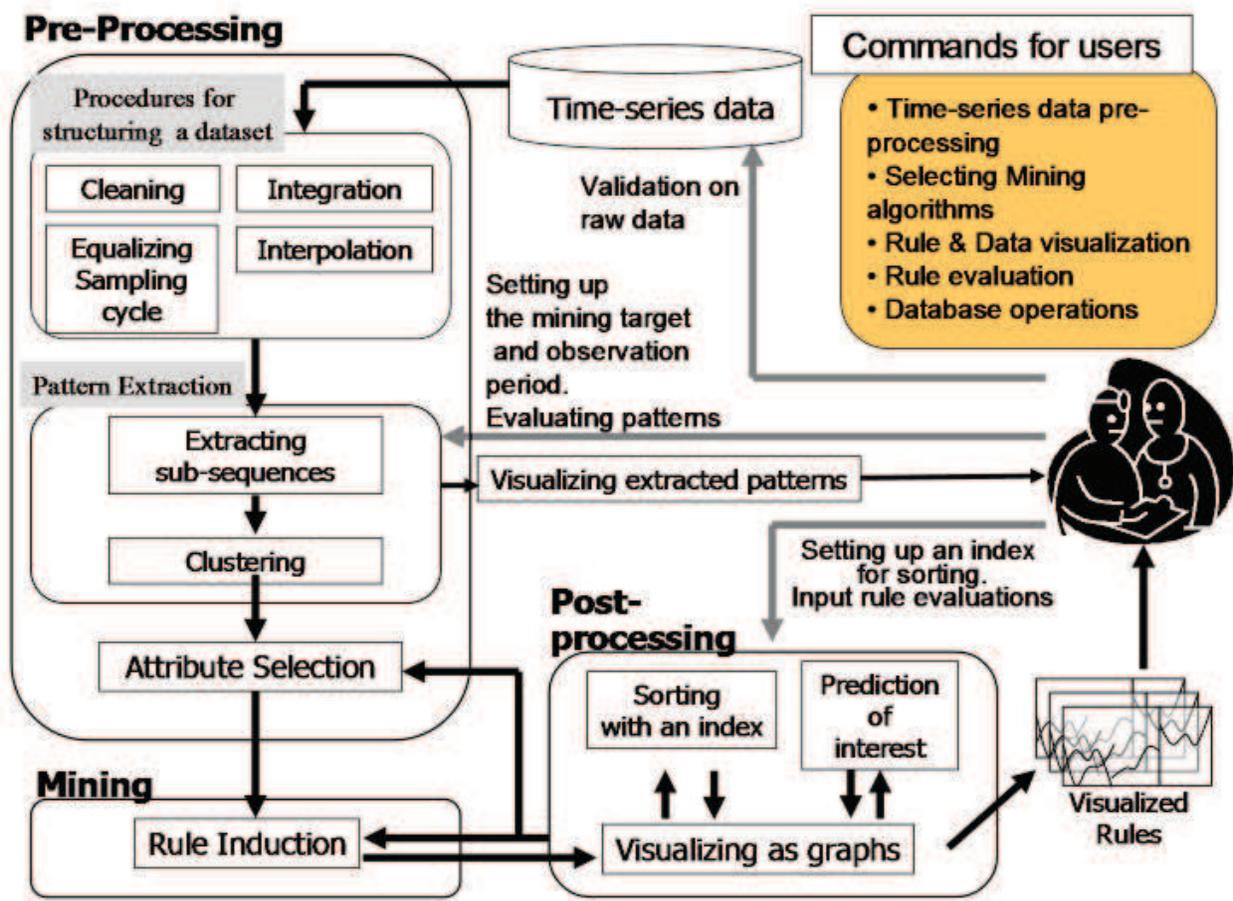


Fig. 1. A system flow view of the integrated time-series data mining environment.

3. Active evaluation and interaction for domain experts

Since we have standardized input/output data formats, data miners and domain experts can execute different algorithms/methods in each procedure seamlessly. They can execute these procedures on graphical human-system interfaces, discussing each other. Beside, system developers can connect new or improved method for a procedure separately. Only following input/output data formats, system developers can also connect a complex sub-system, which selects a proper algorithm/method to the procedure before executing it. If an algorithm/method lacks for a procedure, they are only needed to develop its wrapper to connect the procedure, because each procedure assumes plug-in modules in this environment. To implement the environment, we have analyzed time-series data mining frameworks. Then we have identified procedures for pattern extraction as data pre-processing, rule induction as mining, and evaluation of rules with visualized rule as post-processing of mined result. The system provides these procedures as commands for users. At the same time, we have designed graphical interfaces, which include data processing, validation for patterns on elemental sequences, and rule visualization as graphs. Fig. 1 shows us a typical system flow of this time-series data mining environment.

3.1 Mining classification rules consisting of temporal patterns

In order to obtain classification rules with temporal patterns in their consequents, we firstly collect temporal data for the objective problem. For the temporal data, we have identified

procedures for temporal data mining as follows:

- Data pre-processing
 - pre-processing for data construction
 - temporal pattern extraction
 - attribute selection
- Mining
 - classification rule induction
- Other database procedures
 - selection with conditions
 - join

As data pre-processing procedures, pre-processing for data construction procedures include data cleaning, equalizing sampling rate, interpolation, and filtering irrelevant data. Since these procedures are almost manual procedures, they strongly depend on given temporal data and a purpose of the mining process. Temporal pattern extraction procedures include determining the period of sub-sequences and finding representative sequences with a clustering algorithm such as K-Means, EM clustering (Liao, 2005) and the temporal pattern extraction method developed by Ohsaki et al. (Ohsaki, Abe & Yamaguchi, 2007). Attribute selection procedures are done by selecting relevant attributes manually or using attribute selection algorithms (Liu & Motoda, 1998). At mining phase, we should choose a proper rule induction algorithm with some criterion. There are so many classification rule induction algorithms such as Version Space (Mitchell, 1982), AQ15 (Michalski, 1986), C4.5 rule (Quinlan, 1993), and any other algorithm. To support this choice, we have developed a tool to construct a proper mining application based on constructive meta-learning called CAMLET (Abe & Yamaguchi, 2004). However, we have taken PART (Frank et al., 1998) implemented in Weka (Witten & Frank, 2000) in the case study to evaluate improvement of our pattern extraction algorithm.

3.2 Prediction for test data with classifying temporal patterns and evaluation with visualizing rules

- Post-processing of mined results
 - predicting classes of test(unknown) data
 - visualizing mined rule
 - rule selection
 - supporting rule evaluation

In order to predict class of a test dataset with learned a classification model, the system should formally predict pattern symbols of the test dataset using some accurate classification learning method L^2 based on the training dataset as shown in Fig. 2.

²Since this classification learning algorithm is not required understandability of the learning model, we can use more complicate but accurate learning algorithms such as neural network and ensemble learning scheme in this process.

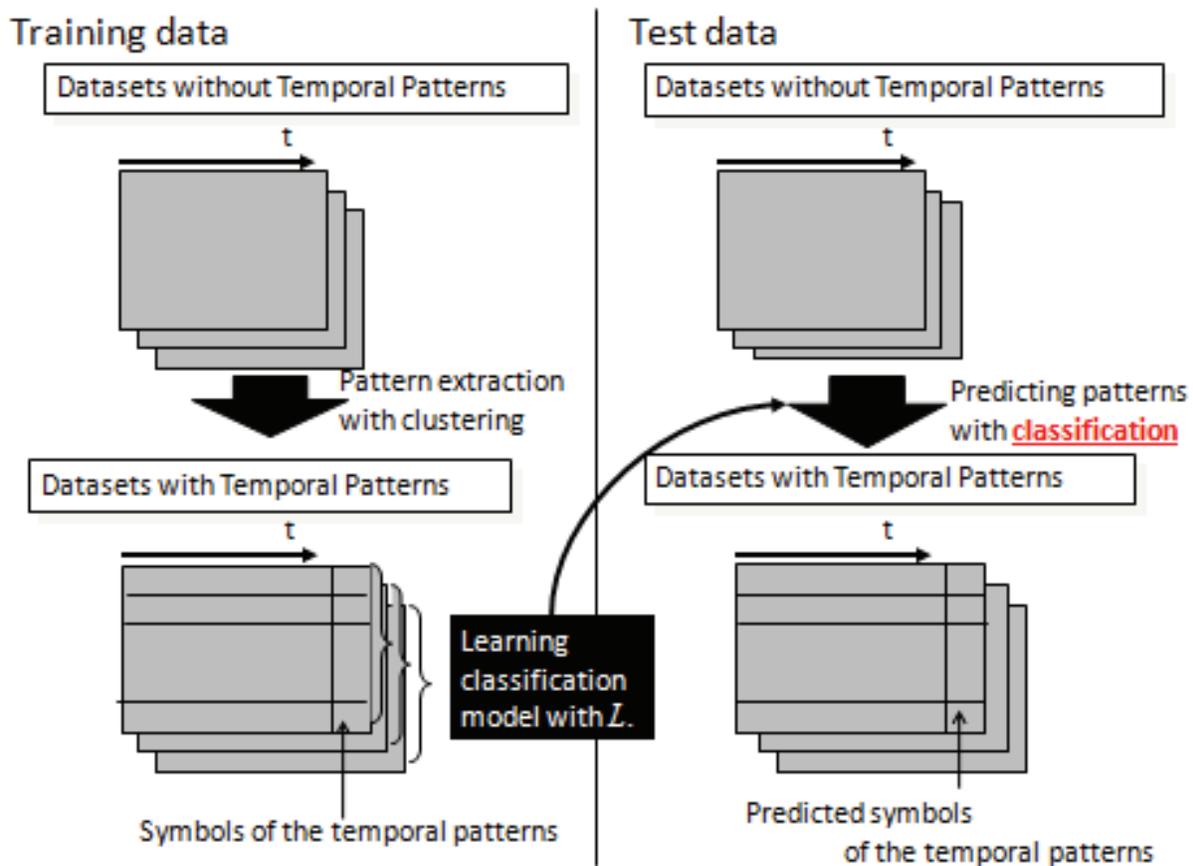


Fig. 2. Temporal pattern prediction phase for test data to predict by using the mined classification rules with temporal patterns.

To validate mined rules correctly, users need readability and ease for understand about mined results. We have taken 39 objective rule evaluation indexes to select mined rules (Ohsaki, Abe, Tsumoto, Yokoi & Yamaguchi, 2007), visualizing and sorting them depended on usersf interest. Although these two procedures are passive support from a viewpoint of the system, we have also identified active system reaction with prediction of user evaluation based on objective rule evaluation indexes and human evaluations.

Other database procedures are used to make target data for a data mining process. Since the environment has been designed based on open architecture, these procedures have been able to develop separately. To connect each procedure, we have only defined input/output data format by using the comma separated value style.

4. Temporal rule mining for japanese stock trading

After implementing the integrated temporal data mining environment described in Section 3, we have done a case study on Japanese stock market database. In this case study, we firstly gathered temporal price data and its trend index values through Kaburobo SDK (Kaburobo, 2004). Then, using the environment, we evaluated the performance of if-then rules based on temporal patterns. Finally, with regarding to the results, we discuss about the availability of our temporal rule mining based on temporal pattern extraction.

Attribute name	Description	
RAW	opening	opening price of the day (O_t)
	high	Highest price of the day (H_t)
	low	Lowest price of the day (L_t)
	closing	Closing price of the day (C_t)
	Volume	Volume of the day (V_t)
TREND INDEXES	Moving Average	Buy: if $SMA_t - LMA_t < 0 \cap SMA_{t+1} - LMA_{t+1} > 0$, Sell: if $SMA_t - LMA_t > 0 \cap SMA_{t+1} - LMA_{t+1} < 0$ Where $SMA_t = (C_t + C_{t+1} + \dots + C_{t+12}) / 13$, and $LMA_t = (C_t + C_{t+1} + \dots + C_{t+25}) / 26$
	Bolinger Band	Buy: if $C_t \geq (MA_t + 2\sigma) \times 0.05$, Sell: if $C_t \leq (MA_t - 2\sigma) \times 0.05$ where $MA_t = (C_t + C_{t+1} + \dots + C_{t+25}) / 26$
	Envelope	Buy: if $C_t \geq MA_t + (MA_t \times 0.05)$, Sell: if $C_t \leq MA_t - (MA_t \times 0.05)$
	HLband	Buy: if $C_t < LowLine_{t-20days}$, Sell: if $C_t > HighLine_{t-20days}$
	MACD	Buy: if $MACD_t - AvgMACD_{t-20days} > 0 \cap MACD_{t+1} - AvgMACD_{t+1-20days} < 0$ Sell: if $MACD_t - AvgMACD_{t-20days} < 0 \cap MACD_{t+1} - AvgMACD_{t+1-20days} > 0$ Where $MACD_t = EMA_{t,0.05} - EMA_{t,26days}$, $EMA_t = EMA_{t-1} + (2/range) \times (C_{t-1} - EMA_{t-1})$
	DMI	Buy: if $PDI_t - MDI_t > 0 \cap PDI_{t+1} - MDI_{t+1} < 0$, Sell: if $PDI_t - MDI_t < 0 \cap PDI_{t+1} - MDI_{t+1} > 0$ Where $PDI_t = \sum_{i=t-25}^t (H_i - L_{i+1}) / \sum_{i=t-25}^t TR_i \times 100$, $MDI_t = \sum_{i=t-25}^t (L_i - H_{i+1}) / \sum_{i=t-25}^t TR_i \times 100$ $TR_t = \max\{(H_t - C_{t-1}), (C_t - L_t), (H_t - L_t)\}$
	volumeRatio	$VR_t = \{(\sum_{i=t-25}^t V_i + \sum_{i=t-26}^t V_i) / (\sum_{i=t-25}^t V_i + \sum_{i=t-26}^t V_i)\} \times 100$
	RSI	$RSI_t = 100 - 100 / (1 + \sum_{i=t-25}^t (C_{i+1} - C_i) / \sum_{i=t-25}^t (C_{i+1} - C_i) + 1)$
	Momentum	$M_t = C_t - C_{t-10}$
	ichimoku1	Buy: if $C_{t-1} < RL_{t-9days} \cap C_t > RL_{t-9days}$, Sell: if $C_{t+1} > RL_{t-9days} \cap C_t < RL_{t-9days}$ Where $RL_{t-9days} = average(\max(H_i) + \min(L_i))$ ($i = t - 8, t - 7, \dots, t$)
	ichimoku2	Buy: if $C_{t-1} < RL_{t-26days} \cap C_t > RL_{t-26days}$, Sell: if $C_{t+1} > RL_{t-26days} \cap C_t < RL_{t-26days}$ Where $RL_{t-26days} = average(\max(H_i) + \min(L_i))$ ($i = t - 25, t - 24, \dots, t$)
	ichimoku3	Buy: if $RL_{(t-2)-26days} < RL_{(t-1)-9days} \cap RL_{(t-1)-26days} > RL_{(t-1)-9days} \cap RL_{(t-1)-26days} < RL_{(t-1)-26days}$ Sell: if $RL_{(t-2)-26days} > RL_{(t-1)-9days} \cap RL_{(t-1)-26days} < RL_{(t-1)-9days} \cap RL_{(t-1)-26days} > RL_{(t-1)-26days}$
	ichimoku4	Buy: if $C_t > AS1_{t-25} \cap C_t > AS2_{t-25}$, Sell: if $C_t < AS1_{t-25} \cap C_t < AS2_{t-25}$ Where $AS1_t = median(RL_{t-3days} - RL_{t-26days})$, $AS2_t = (\max(H_t) - \min(L_t)) / 2$ ($t = t - 51, t - 50, \dots, t$)

Table 1. The description about attributes from Kaburobo SDK.

4.1 Description about temporal datasets

Using Kaburobo SDK, we got four price values, trading volume, and 13 trend index values as shown in Table 1. The daily four price volumes and daily trading volume of each stock are gathered from the Kaburobo SDK as raw values. Then, we set up 13 days as short term range and 26 days as long term range for calculating technical indexes that consider both of short and long terms. Excepting DMI, volume ratio, and momentum, the trend indices are defined as trading signals: buy and sell. The attribute values of these indices are converted from 1.0 to -1.0. Thus, 0 means nothing to do (or hold on the stock) for these attributes.

We obtained temporal data consists of the above mentioned attributes about five financial companies and four telecommunication companies as follows: Credit Saison (Saison), Orix, Mitsubishi Tokyo UFJ Financial Group (MUFG), Mitsui Sumitomo Financial Group (MSFG), Mizuho Financial Group (MizuhoFG), NTT, KDDI, NTT Docomo (NTTdocomo), and Softbank. The period, which we have collected from the temporal stock data, is from 5th January 2006 to 31st May 2006. For each day, we have made decisions as the following: the decision is if the closing value rises 5% within 20 days then ‘buy’, otherwise if the closing value falls 5% within 20 days then ‘sell’, otherwise ‘hold’.

We set these decisions as the class attribute to each target instance. Table 2 shows the class distributions about the nine stocks for the period.

For each gathered temporal data of the nine stocks, the system extracted temporal patterns for each attribute. To extract temporal patterns, we have used K-Means and Gaussian Mixture Model (GMM) clustering optimized with EM algorithm³, which are implemented in Weka. As for the number of extracted patterns for K-Means as k , we set up $k = 4$ for being easy to

³Hereafter, we call this clustering algorithm as “GMM with EM algorithm.”

Finance	buy	sell	Telecom	buy	sell
Saison	37	53	NTT	27	32
Orix	43	40	KDDI	42	39
MUFJFG	0	50	NTTdocomo	19	29
MSFG	6	27	Softbank	23	69
MizuhoFG	38	31			

Table 2. The class distributions of the nine stocks during the five months.

understand the extracted patterns on each technical indexes and their combinations. Then, the symbols of each pattern and the decision of each day joined as each instance of the target dataset.

5. Evaluating temporal pattern prediction by boosted C4.5

In order to predict temporal pattern of each test dataset, we have used Boosted C4.5 (Quinlan, 1996), which is also implemented in Weka. Table 3 and Table 4 show accuracies of temporal pattern prediction using Boosted C4.5 on patterns obtained by each clustering algorithm. These accuracies are averages of 100 times repeated 10-fold cross validation on the 18 datasets of the technical indexes as the attribute of each target dataset.

5.1 Mining results of the nine temporal stock data

In this section, we show accuracies of temporal rule mining with PART on each datasets themselves and cross-stocks.

As shown in Table 5, each rule set predicts the class labels of training dataset itself on each stock. The accuracies of the nine dataset are satisfactory high scores as a classification task.

As for evaluating the accuracy and efficiency of the classification rules with real value temporal patterns, we performed a cross-stock evaluation. In this evaluation, we obtained a rule set from one stock, and apply it to the other stock for predicting class labels; sell, buy, or hold. Table 6 and Table 6 show accuracies (The cross stock evaluation uses different stocks as training dataset and test dataset. Stocks in rows mean training datasets, and columns mean test datasets. As shown in these tables, emphasized numbers go beyond 50%, which means that the mined rules work better than just predicting sell or buy. The result shows the performance of our temporal rules depends on the similarity of trend values rather than the field of each stock.

5.2 Detailed result of the obtained classification rules with temporal patterns

As shown in Table 6 and Table 6, some rule sets predict significant decisions compared to the random prediction. In order to describe the rules more clearly, we present the representative rules in this section.

Fig. 3 shows an example of the classification rules with temporal patterns. These rules are obtained from the training dataset obtained by GMM with EM algorithm temporal pattern extraction for Saison. As shown in Table 6, the rule set of Saison works the best to KDDI as the test dataset.

With regarding Fig. 3, our temporal rule mining system can find out adequate combinations of trend index patterns for each stock. To learn adequate trend index pattern combinations is very costly work for trading beginners. Thus, our temporal rule mining can support traders who want to know the adequate combinations of trend indexes for each stock.

IndexName	Saison	MUFJFG	MSFG	MizuhoFG	Orix	KDDI	NTT	NTTdoc
opening	88.0	83.0	86.0	89.0	83.0	93.0	92.0	
high	84.0	88.0	94.0	87.0	83.0	93.0	91.0	
low	85.0	92.0	90.0	92.0	81.0	93.0	91.0	
closing	86.0	86.0	93.0	91.0	74.0	93.0	92.0	
volume	70.0	79.0	86.0	72.0	71.0	79.0	80.0	
MovingAvg.	96.0	94.9	84.8	88.9	81.8	91.9	62.6	
BollingerBand	94.9	90.9	79.8	93.9	94.9	80.8	100.0	
Envelope	89.9	89.9	93.9	89.9	89.9	85.9	82.8	
HLband	91.9	83.8	90.9	89.9	83.8	87.9	76.8	
MACD	84.8	91.9	77.8	81.8	91.9	76.8	90.9	
DMI	76.8	84.8	88.9	82.8	90.9	85.9	85.9	
volumeRatio	87.9	87.9	91.9	88.9	90.9	91.9	91.9	
RSI	85.9	88.9	85.9	88.9	83.8	87.9	83.8	
Momentum	82.8	85.9	76.8	81.8	86.9	85.9	82.8	
Ichimoku1	67.7	92.9	90.9	86.9	74.7	48.5	87.9	
Ichimoku2	58.6	87.9	77.8	82.8	83.8	58.6	73.7	
Ichimoku3	97.0	97.0	94.9	74.7	100.0	100.0	100.0	
Ichimoku4	78.8	84.8	93.9	89.9	91.9	73.7	93.9	

Table 3. Accuracies (%) of temporal pattern prediction by Boosted C4.5 on the patterns obtained by K-Means.

IndexName	Saison	MUFJFG	MSFG	MizuhoFG	Orix	KDDI	NTT	NTTdoco
opening	88.0	93.0	86.0	89.0	99.0	90.0	90.0	9
high	90.0	88.0	85.0	96.0	93.0	90.0	91.0	9
low	94.0	91.0	92.0	90.0	94.0	92.0	80.0	9
closing	87.0	95.0	83.0	92.0	97.0	93.0	84.0	9
volume	72.0	58.0	77.0	64.0	64.0	71.0	81.0	6
MovingAvg.	49.5	63.6	65.7	54.5	63.6	43.4	42.4	5
BollingerBand	74.7	82.8	80.8	59.6	86.9	48.5	100.0	7
Envelope	85.9	90.9	85.9	60.6	74.7	78.8	57.6	10
HLband	71.7	89.9	84.8	87.9	79.8	68.7	58.6	5
MACD	63.6	51.5	49.5	49.5	58.6	64.6	44.4	5
DMI	55.6	62.6	69.7	45.5	80.8	57.6	38.4	5
volumeRatio	89.9	81.8	92.9	93.9	81.8	84.8	85.9	9
RSI	85.9	88.9	92.9	86.9	80.8	89.9	81.8	8
Momentum	84.8	87.9	81.8	88.9	89.9	85.9	78.8	8
Ichimoku1	47.5	39.4	45.5	47.5	54.5	56.6	60.6	5
Ichimoku2	50.5	46.5	51.5	63.6	58.6	54.5	53.5	4
Ichimoku3	72.7	80.8	75.8	97.0	100.0	100.0	100.0	8
Ichimoku4	82.8	82.8	87.9	94.9	62.6	76.8	78.8	8

Table 4. Accuracies (%) of temporal pattern prediction by Boosted C4.5 on the patterns obtained by GMM with EM algorithm.

Stock Name	K-Means	EM	Stock Name	K-Means	EM
Saison	90.1	88.9	NTT	84.8	90.9
Orix	88.9	84.8	KDDI	86.9	78.8
MUFJFG	90.9	93.9	NTTdocomo	80.8	85.9
MSFG	96.0	90.9	Softbank	93.9	89.9
MizuhoFG	92.9	83.8			

Table 5. Re-substitution accuracies (%) of the rule sets obtained with the two temporal pattern extraction with K-Means and GMM with EM algorithm.

The shapes of the visualized temporal patterns and their combination are also useful for supporting knowledge discovery in medical data. As described in (Abe et al., 2007), a physician could find interesting temporal patterns of ALT (alanine transaminase) and RBC (Red Blood-cell Count) related to the good result of the INF (interferon) treatment.

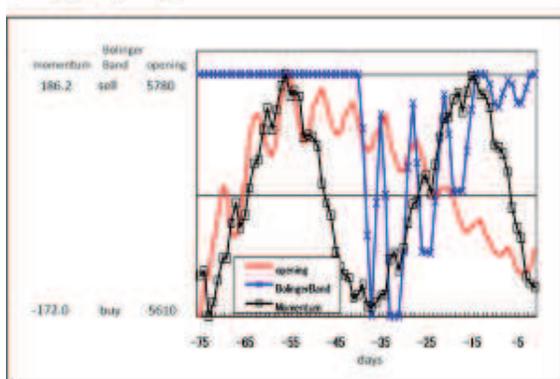
6. Conclusion

In this chapter, we described the temporal classification rule mining framework that combines time-series data pre-processing, mining, post-processing for mined results, and other database operators to validate data and results of every phase.

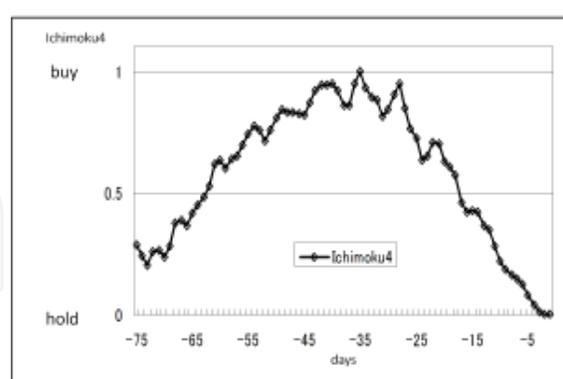
By using the implementations and experimental results, we present the following merits achieved by the classification rules with considering temporal patterns of the target attributes:

- Finding different interesting aspects of the decisions/results
- Finding important temporal patterns and attributes for the decision at the same time

In order to utilize this framework with proper methods in each step, we expect that the readers may construct their own temporal classification rule mining systems for their sequential data by combining other temporal pattern extraction methods and classification rules.



(a) Example for 'buy'



(b) Example for 'sell'

Fig. 3. An example of rule for 'buy' and rule for 'sell'.

Training\Test	Saison	MUFJFG	MSFG	MizuhoFG	Orix	NTT	KDDI
Saison		44.4	28.3	31.3	40.4	29.3	35.4
MUFJFG	46.5		44.4	30.3	42.4	32.3	39.4
MSFG	44.4	24.2		38.4	31.3	28.3	27.3
MizuhoFG	46.5	31.3	33.3		29.3	22.2	20.2
Orix	38.4	50.5	27.3	31.3		32.3	39.4
NTT	14.1	50.5	27.3	31.3	14.1		39.4
KDDI	12.1	44.4	56.6	27.3	31.3	41.4	
NTTdocomo	26.3	40.4	52.5	33.3	23.2	30.3	20.2
Softbank	44.4	28.3	18.2	45.5	34.3	40.4	30.3

(a)

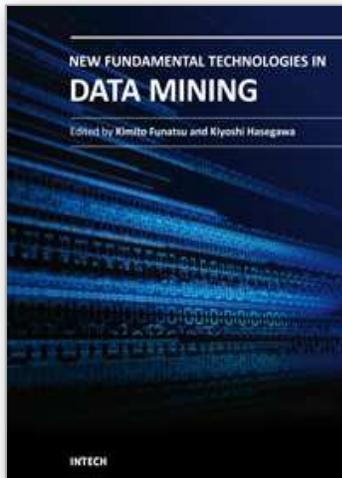
Training\Test	Saison	MUFJFG	MSFG	MizuhoFG	Orix	NTT	KDDI
Saison		46.5	28.3	31.3	38.4	51.5	65.5
MUFJFG	31.3		51.5	31.3	38.4	29.3	41.4
MSFG	23.2	58.6		34.3	31.3	43.4	32.3
MizuhoFG	35.4	31.3	34.3		31.3	42.4	38.4
Orix	41.4	29.3	39.4	34.3		37.4	21.2
NTT	41.4	21.2	20.2	42.4	44.4		33.3
KDDI	61.6	59.6	50.5	28.3	27.3	42.4	
NTTdocomo	27.3	42.4	29.3	52.5	25.3	30.3	19.2
Softbank	52.5	45.5	27.3	31.3	41.4	33.3	43.4

(b)

Table 6. a) Accuracies (%) of cross stock evaluation with the temporal patterns obtained by K-Means. b) Accuracies (%) of cross stock evaluation with the temporal patterns obtained by GMM with EM algorithm.

7. References

- Abe, H. & Yamaguchi, T. (2004). Constructive meta-learning with machine learning method repositories, *Proc. of the 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE 2004*, LNAI 3029, pp. 502–511.
- Abe, H., Yokoi, H., Ohsaki, M. & Yamaguchi, T. (2007). Developing an integrated time-series data mining environment for medical data mining, *ICDMW '07: Proceedings of the Seventh IEEE International Conference on Data Mining Workshops*, IEEE Computer Society, Washington, DC, USA, pp. 127–132.
- Akaike, H. (1969). Fitting autoregressive models for prediction, 21(1): 243–247.
- Berndt, D. J. & Clifford, J. (1996). Finding patterns in time series: a dynamic programming approach, pp. 229–248.
- Das, G., King-Ip, L., Heikki, M., Renganathan, G. & Smyth, P. (1998). Rule discovery from time series, *Proc. of International Conference on Knowledge Discovery and Data Mining*, pp. 16–22.
- Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery: an overview, pp. 1–34.
- Frank, E., Wang, Y., Inglis, S., Holmes, G. & Witten, I. H. (1998). Using model trees for classification, *Machine Learning* 32(1): 63–76.
- Hirano, S. & Tsumoto, S. (2002). Multiscale comparison of temporal patterns in time-series medical databases, *PKDD '02: Proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discovery*, Springer-Verlag, London, UK, pp. 188–199.
- Kaburobo (2004). <http://www.kaburobo.jp/>.
- Keogh, E., Chu, S., Hart, D. & Pazzani, M. (2003). Segmenting time series: A survey and novel approach, *an Edited Volume, Data mining in Time Series Databases.*, World Scientific, pp. 1–22.
- Liao, T. W. (2005). Clustering of time series data: a survey, *Pattern Recognition* 38: 1857–1874.
- Liu, H. & Motoda, H. (1998). *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers, Norwell, MA, USA.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11(7): 674–693.
- Mitchell, T. M. (1982). Generalization as search, *Artificial Intelligence* 18(2): 203–226.
- Ohsaki, M., Abe, H., Tsumoto, S., Yokoi, H. & Yamaguchi, T. (2007). Evaluation of rule interestingness measures in medical knowledge discovery in databases, *Artif. Intell. Med.* 41(3): 177–196.
- Ohsaki, M., Abe, H. & Yamaguchi, T. (2007). Numerical time-series pattern extraction based on irregular piecewise aggregate approximation and gradient specification, *New Gen. Comput.* 25(3): 213–222.
- Ohsaki, M., Kitaguchi, S., Kume, S., Yokoi, H. & Yamaguchi, T. (2004). Evaluation of rule interestingness measures with a clinical dataset on hepatitis, *Proceedings of ECML/PKDD 2004*, LNAI3202, pp. 362–373.
- Quinlan, J. R. (1993). *Programs for Machine Learning*, Morgan Kaufmann Publishers.
- Quinlan, J. R. (1996). Bagging, boosting, and c4.5, *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, AAAI Press, pp. 725–730.
- Witten, I. H. & Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann.
- Wong, R. C.-W. & Fu, A. W.-C. (2006). Mining top-k frequent itemsets from data streams, *Data Min. Knowl. Discov.* 13(2): 193–217.



New Fundamental Technologies in Data Mining

Edited by Prof. Kimito Funatsu

ISBN 978-953-307-547-1

Hard cover, 584 pages

Publisher InTech

Published online 21, January, 2011

Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hidenao Abe (2011). Development of a Classification Rule Mining Framework by Using Temporal Pattern Extraction, New Fundamental Technologies in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-547-1, InTech, Available from: <http://www.intechopen.com/books/new-fundamental-technologies-in-data-mining/development-of-a-classification-rule-mining-framework-by-using-temporal-pattern-extraction>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen