

Discovery of Temporal Knowledge in Medical Time-Series Databases Using Moving Average, Multiscale Matching, and Rule Induction

Shusaku Tsumoto

Department of Medicine Informatics, Shimane Medical University, School of Medicine
89-1 Enya-cho Izumo City, Shimane 693-8501 Japan
tsumoto@computer.org

Abstract. Since hospital information systems have been introduced in large hospitals, a large amount of data, including laboratory examinations, have been stored as temporal databases. The characteristics of these temporal databases are: (1) Each record are inhomogeneous with respect to time-series, including short-term effects and long-term effects. (2) Each record has more than 1000 attributes when a patient is followed for more than one year. (3) When a patient is admitted for a long time, a large amount of data is stored in a very short term. Even medical experts cannot deal with these large databases, the interest in mining some useful information from the data are growing. In this paper, we introduce a combination of extended moving average method, multiscale matching and rule induction method to discover new knowledge in medical temporal databases. This method was applied to a medical dataset, the results of which show that interesting knowledge is discovered from each database.

1 Introduction

Since hospital information systems have been introduced in large hospitals, a large amount of data, including laboratory examinations, have been stored as temporal databases [15]. For example, in a university hospital, where more than 1000 patients visit from Monday to Friday, a database system stores more than 1 GB numerical data of laboratory examinations. Thus, it is highly expected that data mining methods will find interesting patterns from databases because medical experts cannot deal with those large amount of data. The characteristics of these temporal databases are: (1) Each record are inhomogeneous with respect to time-series, including short-term effects and long-term effects. (2) Each record has more than 1000 attributes when a patient is followed for more than one year. (3) When a patient is admitted for a long time, a large amount of data is stored in a very short term. Even medical experts cannot deal with these large temporal databases, the interest in mining some useful information from the data are growing.

In this paper, we introduce a rule discovery method, combined with extended moving average method, multiscale matching for qualitative trend to discover

new knowledge in medical temporal databases. In this system, extended moving average method and multi-scale matching are used for preprocessing, to deal with irregularity of each temporal data. Using several parameters for time-scaling, given by users, this moving average method generates a new database for each time scale with summarized attributes. For matching time sequences, multiscale matching was applied. Then, rule induction method is applied to each new database with summarized attributes. This method was applied to two medical datasets, the results of which show that interesting knowledge is discovered from each database.

This paper is organized as follows. Section 2 introduces the definition of probabilistic rules. Section 3 discusses the characteristics of temporal databases in hospital information systems. Section 4 presents extended moving average method. Section 5 introduces second preprocessing methods to extract qualitative trend and rule discovery method with qualitative trend. Section 6 shows experimental results. Section 7 gives a brief discussion of the total method. Finally, Section 8 concludes this paper.

2 Probabilistic Rules and Conditional Probabilities

Before discussing temporal knowledge discovery, we first discuss the characteristics of probabilistic rules. In this section, we use the following notations introduced by Grzymala-Busse and Skowron [11], which are based on rough set theory [10].

Let U denote a nonempty, finite set called the universe and A denote a nonempty, finite set of attributes, i.e., $a : U \rightarrow V_a$ for $a \in A$, where V_a is called the domain of a , respectively. Then, a decision table is defined as an information system, $A = (U, A \cup \{d\})$. The atomic formulae over $B \subseteq A \cup \{d\}$ and V are expressions of the form $[a = v]$, called descriptors over B , where $a \in B$ and $v \in V_a$. The set $F(B, V)$ of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation. For each $f \in F(B, V)$, f_A denote the meaning of f in A , i.e., the set of all objects in U with property f , defined inductively as follows.

1. If f is of the form $[a = v]$ then, $f_A = \{s \in U | a(s) = v\}$
2. $(f \wedge g)_A = f_A \cap g_A$; $(f \vee g)_A = f_A \cup g_A$; $(\neg f)_A = U - f_A$

By the use of the framework above, classification accuracy and coverage, or true positive rate is defined as follows.

Let R and D denote a formula in $F(B, V)$ and a set of objects which belong to a decision d . Classification accuracy and coverage (true positive rate) for $R \rightarrow d$ is defined as:

$$\alpha_R(D) = \frac{|R_A \cap D|}{|R_A|} (= P(D|R)), \text{ and } \kappa_R(D) = \frac{|R_A \cap D|}{|D|} (= P(R|D)),$$

where $|S|$, $\alpha_R(D) (= P(D|R))$, $\kappa_R(D) (= P(R|D))$ and $P(S)$ denote the cardinality of a set S , a classification accuracy of R as to classification of D and coverage (a true positive rate of R to D), and probability of S , respectively.

By the use of accuracy and coverage, a probabilistic rule is defined as:

$$R \xrightarrow{\alpha, \kappa} d \quad \text{s.t.} \quad R = \bigwedge_j [a_j = v_k], \alpha_R(D)(= P(D|R)) \geq \delta_\alpha \\ \text{and } \kappa_R(D)(= P(R|D)) \geq \delta_\kappa,$$

For further information about these probabilistic rules, reader may refer to [14].

3 Temporal Databases in Hospital Information Systems

Since incorporating temporal aspects into databases is still an ongoing research issue in database area [1], temporal data are stored as a table in hospital information systems(H.I.S.). Table 1 shows a typical example of medical data, which is retrieved from H.I.S. The first column denotes the ID number of each patient, and the second one denotes the date when the datasets in this row is examined. Each row with the same ID number describes the results of laboratory examinations, which were taken on the date in the second column. For example, the second row shows the data of the patient ID 1 on 04/19/1986. This simple database show the following characteristics of medical temporal database:

(1) The Number of Attributes Are Too Many. Even though the dataset of a patient focuses on the transition of each examination (attribute), it would be difficult to see its trend when the patient is followed for a long time. If one wants to see the long-term interaction between attributes, it would be almost impossible. In order to solve this problems, most of H.I.S. systems provide several graphical interfaces to capture temporal trends [15]. However, the interactions among more than three attributes are difficult to be studied even if visualization interfaces are used.

(2) Irregularity of Temporal Intervals. Temporal intervals are irregular. Although most of the patients will come to the hospital every two weeks or one month, physicians may not make laboratory tests at each time. When a patient has a acute fit or suffers from acute diseases, such as pneumonia, laboratory examinations will be made every one to three days. On the other hand, when his/her status is stable, these test may not be made for a long time. Patient ID 1 is a typical example. Between 04/30 and 05/08/1986, he suffered from a pneumonia and was admitted to a hospital. Then, during the therapeutic procedure, laboratory tests were made every a few days. On the other hand, when he was stable, such tests were ordered every one or two year.

(3) Missing Values. In addition to irregularity of temporal intervals, datasets have many missing values. Even though medical experts will make laboratory examinations, they may not take the same tests in each instant. Patient ID 1 in Table 1 is a typical example. On 05/06/1986, medical physician selected a specific test to confirm his diagnosis. So, he will not choose other tests. On 01/09/1989, he focused only on GOT, not other tests. In this way, missing values will be observed very often in clinical situations.

Table 1. An example of temporal database

ID	Date	GOT	GPT	LDH	γ -GTP	TP	edema	...
1	19860419	24	12	152	63	7.5	-	...
1	19860430	25	12	162	76	7.9	+	...
1	19860502	22	8	144	68	7.0	+	...
1	19860506							...
1	19860508	22	13	156	66	7.6	-	...
1	19880826	23	17	142	89	7.7	-	...
1	19890109	32					-	...
1	19910304	20	15	369	139	6.9	+	...
2	19810511	20	15	369	139	6.9	-	...
2	19810713	22	14	177	49	7.9	-	...
2	19880826	23	17	142	89	7.7	-	...
2	19890109	32					-	...
...								

These characteristics have already been discussed in KDD area [5]. However, in real-world domains, especially domains in which follow-up studies are crucial, such as medical domains, these ill-posed situations will be distinguished. If one wants to describe each patient (record) as one row, then each row have too many attributes, which depends on how many times laboratory examinations are made for each patient. It is notable that although the above discussions are made according to the medical situations, similar situations may occur in other domains with long-term follow-up studies.

4 Extended Moving Average Methods

4.1 Moving Average Methods

Averaging mean methods have been introduced in statistical analysis [6]. Temporal data often suffers from noise, which will be observed as a spike or sharp wave during a very short period, typically at one instant. Averaging mean methods remove such an incidental effect and make temporal sequences smoother.

With one parameter w , called *window*, moving average \hat{y}_w is defined as follows:

$$\hat{y}_w = \sum_{j=1}^w y_j.$$

For example, in the case of GOT of patient ID 1, y_5 is calculated as: $\hat{y}_5 = (24 + 25 + 22 + 22 + 22)/5 = 23.0$. It is easy to see that \hat{y}_w will remove the noise effect which continue less than w points.

The advantage of moving average method is that it enables to remove the noise effect when inputs are given periodically [6]. For example, when some tests are measured every several days¹, the moving average method is useful

¹ This condition guarantees that measurement is approximately continuous

to remove the noise and to extract periodical domains. However, in real-world domains, inputs are not always periodical, as shown in Table 1. Thus, when applied time-series are irregular or discrete, ordinary moving average methods are powerless. Another disadvantage of this method is that it cannot be applicable to categorical attributes. In the case of numerical attributes, average can be used as a summarized statistic. On the other hand, such average cannot be defined for categorical attributes.

Thus, we introduce the extended averaging method to solve these two problems in the subsequent subsections.

4.2 Extended Moving Average for Continuous Attributes

In this extension, we first focus on how moving average methods remove noise. The key idea is that a window parameter w is closely related with periodicity. If w is larger, then the periodical behavior whose time-constant is lower than w will be removed. Usually, a spike by noise is observed as a single event and this effect will be removed when w is taken as a large value. Thus, the choice of w separates different kinds of time-constant behavior in each attribute and in the extreme case when w is equal to total number of temporal events, all the temporal behavior will be removed. We refer to this extreme case as $w = \infty$.

The extended moving average method is executed as follows: first calculates y_∞ for an attribute y . Second, the method outputs its maximum and minimum values. Then, according to the selected values for w , a set of sequence $\{y_w(i)\}$ for each record is calculated. For example, if $\{w\}$ is equal to $\{10 \text{ years}, 5 \text{ years}, 1 \text{ year}, 3 \text{ months}, 2 \text{ weeks}\}$, then for each element in $\{w\}$, the method uses the time-stamp attribute for calculation of each $\{y_w(i)\}$ in order to deal with irregularities.

In the case of Table 1, when w is taken as 1 year, all the rows are aggregated into several components as shown in Table 2. From this aggregation, a sequence y_w for each attribute is calculated as in Table 3.

Table 2. Aggregation for $w = 1$ (year)

ID	Date	GOT	GPT	LDH	γ -GTP	TP	edema	...
1	19860419	24	12	152	63	7.5	-	...
1	19860430	25	12	162	76	7.9	+	...
1	19860502	22	8	144	68	7.0	+	...
1	19860506							...
1	19860508	22	13	156	66	7.6	-	...
1	19880826	23	17	142	89	7.7	-	...
1	19890109	32					-	...
1	19910304	20	15	369	139	6.9	+	...

...

Table 3. Moving average for $w = 1$ (year)

ID	Period	GOT	GPT	LDH	γ -GTP	TP	edema	...
1	1	23.25	11.25	153.5	68.25	7.5	?	...
1	2	23	17	142	89	7.7	?	...
1	3	32					?	...
1	4						?	...
1	5	20	15	369	139	6.9	?	...
1	∞	24	12.83	187.5	83.5	7.43	?	...
...								

The selection of w can be automated. The simplest way to calculate w is to use the power of natural number, such as 2. For example, we can use 2^n as the window length: 2, 4, 8, 16,..... Using this scale, two weeks, three months, one year correspond to $16 = 2^4$, $64 = 2^6$, $256 = 2^8$.

4.3 Categorical Attributes

One of the disadvantages of moving average method is that it cannot deal with categorical attributes. To solve this problem, we will classify categorical attributes into three types, whose information should be given by users. The first type is *constant*, which will not change during the follow-up period. The second type is *ranking*, which is used to rank the status of a patient. The third type is *variable*, which will change temporally, but ranking is not useful. For the first type, extended moving average method will not be applied. For the second one, integer will be assigned to each rank and extended moving average method for continuous attributes is applied. On the other hand, for the third one, the temporal behavior of attributes is transformed into statistics as follows.

First, the occurrence of each category (value) is counted for each window. For example, in Table 2, *edema* is a binary attribute and variable. In the first window, an attribute *edema* takes $\{-, +, +, -\}$.² So, the occurrence of $-$ and $+$ are 2 and 2, respectively. Then, each conditional probability will be calculated. In the above example, probabilities are equal to $p(-|w_1) = 2/4$ and $p(+|w_1) = 2/4$. Finally, for each probability, a new attribute is appended to the table (Table 4).

Summary of Extended Moving Average. All the process of extended moving average is used to construct a new table for each window parameter as the first preprocessing. Then, second preprocessing method will be applied to newly generated tables. The first preprocessing method is summarized as shown in Fig. 1.

² Missing values are ignored for counting.

Table 4. Final table with moving average for $w = 1$ (year)

ID	Period	GOT	GPT	LDH	γ -GTP	TP	edema(+)	edema(-)	...
1	1	23.25	11.25	153.5	68.25	7.5	0.5	0.5	...
1	2	23	17	142	89	7.7	0.0	1.0	...
1	3	32					0.0	1.0	...
1	4						0.0	1.0	...
1	5	20	15	369	139	6.9	1.0	0.0	...
1	∞	24	12.83	187.5	83.5	7.43	0.43	0.57	...
...									

1. Repeat for each w in List L_w ,
 - a) Select an attribute in a List L_a ;
 - i. If an attribute is numerical, then calculate moving average for w ;
 - ii. If an attribute is constant, then break;
 - iii. If an attribute is rank, then assign integer to each ranking; calculate moving average for w ;
 - iv. If an attribute is variable, calculate accuracy and coverage of each category;
 - b) If L_a is not empty, goto (a).
 - c) Construct a new table with each moving average.
2. Construct a table for $w = \infty$.

Fig. 1. First preprocessing method

5 Second Preprocessing and Rule Discovery

5.1 Summarizing Temporal Sequences

From the data table after processing extended moving average methods, several preprocessing methods may be applied in order for users to detect the temporal trends in each attribute. One way is discretization of time-series by clustering introduced by Das [4]. This method transforms time-series into symbols representing qualitative trends by using a similarity measure. Then, time-series data is represented as a symbolic sequence. After this preprocessing, rule discovery method is applied to this sequential data. Another way is to find auto-regression equations from the sequence of averaging means. Then, these quantitative equations can be directly used to extract knowledge or their qualitative interpretation may be used and rule discovery [3], other machine learning methods [7], or rough set method [12] can be applied to extract qualitative knowledge.

In this research, we adopt two modes and transforms databases into two forms: one mode is applying temporal abstraction method [9] with multiscale matching [8] as second preprocessing and transforms all continuous attributes into temporal sequences. The other mode is applying rule discovery to the data after the first preprocessing without second one. The reason why we adopted these two mode is that we focus not only on temporal behavior of each attribute,

but also on association among several attributes. Although Miksch's method [9] and Das's approach [4] are very efficient to extract knowledge about transition, they cannot focus on association between attributes in an efficient way. For the latter purpose, much simpler rule discovery algorithm are preferred.

5.2 Continuous Attributes and Qualitative Trend

To characterize the deviation and temporal change of continuous attributes, we introduce standardization of continuous attributes. For this, we only needs the total average \hat{y}_∞ and its standardization σ_∞ . With these parameters, standardized value is obtained as:

$$z_w = \frac{y_w - \hat{y}_\infty}{\sigma_\infty}.$$

The reason why standardization is introduced is that it makes comparison between continuous attributes much easier and clearer, especially, statistic theory guarantees that the coefficients of a auto-regression equation can be compared with those of another equation [6].

After calculating the standardized values, an extraction algorithm for qualitative trends is applied [9] with multiscale matching briefly shown in the next subsection.

This method is processed as follows: First, this method uses data smoothing with window parameters. Secondly, smoothed values for each attributes are classified into seven categories given as domain knowledge about laboratory test values: extremely low, substantially low, slightly low, normal range, slightly high, substantially high, and extremely high. With these categories, qualitative trends are calculated and classified into the following ten categories by using guideline rules: decrease too fast(A1), normal decrease(A2), decrease too slow(A3), zero change(ZA), dangerous increase(C), increase too fast(B1), normal increase(B2), increase too slow(B3), dangerous decrease(D). For matching temporal sequences with guideline rules, multiscale matching method is applied. For example, if the value of some laboratory tests change from substantially high to normal range within a very short time, the qualitative trend will be classified into A1(decrease too fast). For further information, please refer to [9].

5.3 Multiscale Matching

Multiscale matching is based on two basic ideas: the first one is to use the curvature of the curve to detect the points of inflection. The second idea is to use the scale factor to calculate the curvature of the smoothed curve [8]. The curvature is given as:

$$c(t) = \frac{y''}{(1 + (y')^2)^{3/2}},$$

where $y' = dy/dt$ and $y'' = d^2y/dt^2$. In order to compute the curvature of the curve at varying levels of detail, function y is convolved with a one-dimensional

Gaussian kernel $g(t, \sigma)$ of the width (scaling factor) σ :

$$g(t, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2}.$$

$Y(t, \sigma)$, the convolution of $y(t)$ is defined as:

$$Y(t, \sigma) = y(t) \otimes g(t, \sigma) = \int_{-\infty}^{\infty} y(u) \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} du.$$

According to the characteristics of the convolution, the derivative and the second derivative is calculated as:

$$Y'(t, \sigma) = y(t) \otimes \frac{\partial g(t, \sigma)}{\partial t} \quad \text{and} \quad Y''(t, \sigma) = y(t) \otimes \frac{\partial^2 g(t, \sigma)}{\partial t^2}$$

Using $Y(t, \sigma)$, $Y'(t, \sigma)$ and $Y''(t, \sigma)$, we can calculate the curvature of a given curve for each value of σ within one window w :

$$c(t) = \frac{y(t) \otimes \frac{\partial^2 g(t, \sigma)}{\partial t^2}}{\left(1 + \left(y(t) \otimes \frac{\partial g(t, \sigma)}{\partial t}\right)^2\right)^{3/2}}$$

This gives a sequence of the value of curvature for each time series. If two time series sequence is similar with respect to temporal change, two sequences of curvature will be similar. Furthermore, since we calculate the curvature for each scaling factor, we can compare between these sequences from the local level to global level. For further information, please refer to [8] and [9].

5.4 Rule Discovery Algorithm

For rule discovery, a simple rule induction algorithm discussed in [13] is applied, where continuous attributes are transformed into categorical attributes with a cut-off point. As discussed in Sect. 3, moving average method will remove the temporal effect shorter than a window parameter. Thus, $w = \infty$ will remove all the temporal effect, so this moving average can be viewed as data without any temporal characteristics. If rule discovery is applied to this data, it will generate rules which represents non-temporal association between attributes. In this way, data after processing w -moving average is used to discover association with w or longer time-effect. Ideally, from $w = \infty$ down to $w = 1$, we decompose all the independent time-effect associations between attributes. However, the time-constant in which users are interested will be limited and the moving average method shown in Sect. 3 uses a set of w given by users. Thus, application of rule discovery to each table will generate a sequence of temporal associations between attributes. If some temporal associations will be different from associations with $w = \infty$, then these specific relations will be related with a new discovery.

5.5 Summary of Second Preprocessing and Rule Discovery

Second preprocessing method and rule discovery are summarized as shown in Fig. 2.

1. Calculate \hat{y}_∞ and σ_∞ from the table of $w = \infty$;
2. Repeat for each w in List L_w ;
(w is sorted in a descending order.)
 - a) Select a table of w : T_w ;
 - i. Standardize continuous and ranking attributes;
 - ii. Calculate qualitative trends for continuous and ranking attributes with multiscale matching;
 - iii. Construct a new table for qualitative trends;
 - iv. Apply rule discovery method for temporal sequences;
 - b) Apply rule induction methods to the original table T_w ;

Fig. 2. Second preprocessing and rule discovery

6 Experimental Results: Discovered Results in CVD

The above rule discovery system was applied to a clinical database on cerebrovascular diseases (CVD), which has 2610 records, described by 12 classes. Each record followed up at least 10 years and the averaged number of attributes are 2715. A list of w , $\{w\}$ was set to $\{10 \text{ years}, 5 \text{ years}, 1 \text{ year}, 3 \text{ months}, 2 \text{ weeks}\}$ and thresholds, δ_α and δ_κ were set to 0.60 and 0.30, respectively. One of the most important problems in CVD is whether CVD patients will suffer from mental disorders or dementia and how long it takes each patient to reach the status of dementia.

6.1 Non-temporal Knowledge

Concerning the database on CVD, several interesting rules are derived. The most interesting results are the following positive and negative rules for thalamus hemorrhage:

$$\begin{aligned}
 & [Sex = Female] \wedge [Hemiparesis = Left] \wedge [LOC : positive] \rightarrow Thalamus \\
 & \quad (accuracy : 0.62, coverage : 0.33), \\
 & [Risk : Hypertension] \wedge [Sensory = no] \rightarrow Putamen \\
 & \quad (accuracy : 0.65, coverage : 0.43),
 \end{aligned}$$

Interestingly, LOC(loss of consciousness) under the condition of $[Sex = Female] \wedge [Hemiparesis = Left]$ is an important factor to diagnose thalamic damage. In this domain, any strong correlations between these attributes and others, like MND, have not been found yet. It will be our future work to find what factor will be behind these rules. However, these rules do not include the relations between dementia and brain functions.

Short-Term Effect. As short-term rules, the following interesting rules are discovered:

$$\begin{aligned} & [Gastro : A1] \wedge [Quadriceps : A1] \rightarrow [Dementia : A2] \\ & \quad (accuracy : 0.71, coverage : 0.31, w = 3(months)), \\ & [Gastro : D] \wedge [TA : D] \rightarrow [Dementia : A2] \\ & \quad (accuracy : 0.74, coverage : 0.32, w = 3(months)). \end{aligned}$$

These rules suggests that the rapid decrease of muscle power in the lower extremities are weakly related with the appearance of dementia. However, these knowledge has never been reported and further investigation is required for interpretation.

Long-Term Effect. As long-term rules, the following interesting rules are discovered:

$$\begin{aligned} & [JointPosition : A3] \wedge [Quadriceps : A3] \rightarrow [Dementia : A3] \\ & \quad (accuracy : 0.61, coverage : 0.35, w = 1(year)), \\ & [Gastro : A3] \wedge [Vibration : A3] \rightarrow [Dementia : A3] \\ & \quad (accuracy : 0.87, coverage : 0.33, w = 1(year)). \end{aligned}$$

These rules suggests that combination of the decrease of muscle power in the lower extremities and the increase of sensory disturbance are weakly related with the appearance of dementia. However, these knowledge has neither been reported and further investigation is required for interpretation.

7 Discussion

This paper introduces combination of extended moving average methods as first preprocessing, extraction of qualitative trend as second preprocessing and rule discovery. As discussed in Sects. 3 and 4, this approach is inspired by rule discovery in time series introduced by Das [4]. However, the main differences between Das's approach and our approach are the following.

1. For smoothing data, extended moving average method is introduced.
2. The system incorporates domain knowledge about a continuous attribute to detect its qualitative trend.
3. Qualitative trend are calculated for each time-constant.
4. Rules are discovered with respect to not only associations between qualitative trends but also non-temporal associations.

Using these methods, the system discovered several interesting patterns in a clinical database of different time constant.

The disadvantage of this approach is that the program is not good at extracting periodical behavior of disease processes, or the recurrence of some diseases because the qualitative trends do not support the detection of cycles in temporal behavior of attributes. For these periodical processes, auto-regressive function analysis is much more useful [6]. It will be our future work to extend our approach so that it can deal with periodicity more clearly.

8 Conclusion

In this paper, we introduce a combination of extended moving average method, multiscale matching and rule induction method, to discover new knowledge in temporal databases. In the system, extended moving average method are used for preprocessing, to deal with irregularity of each temporal data. Using several parameters for time-scaling with multiscale matching, given by users, this moving average method generates a new database for each time scale with summarized attributes. Then, rule induction method is applied to each new database with summarized attributes. This method was applied to two medical datasets, the results of which show that interesting knowledge is discovered from each database.

References

1. Abiteboul, S., Hull, R., and Vianu, V. *Foundations of Databases*, Addison-Wesley, New York, 1995.
2. Adams, R.D. and Victor, M. *Principles of Neurology*, 5th edition, McGraw-Hill, NY, 1993.
3. Agrawal, R., Imielinski, T., and Swami, A., Mining association rules between sets of items in large databases, in *Proceedings of the 1993 International Conference on Management of Data (SIGMOD 93)*, pp. 207-216, 1993.
4. Das, G., Lin, K.I., Mannila, H., Renganathan, G. and Smyth, P. Rule discovery from time series. In: *Proceedings of Fourth International Conference on Knowledge Discovery and Data Mining*, pp.16-22, 1998.
5. Fayyad, U.M., et al.(eds.), *Advances in Knowledge Discovery and Data Mining*, AAAI Press, 1996.
6. Hamilton, J.D. *Time Series Analysis*, Princeton University Press, 1994.
7. Langley, P. *Elements of Machine Learning*, Morgan Kaufmann, CA, 1996.
8. Mokhtarian, F. and Mackworth, A. Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes. *IEEE Trans. Pattern. Anal. Machine Intell.*, **PAMI-8**, pp.34-43, 1986.
9. Miksch, S., Horn, W., Popow, C., and Paky, F. Utilizing temporal data abstraction for data validation and therapy planning for artificially ventilated newborn infants. *Artificial Intelligence in Medicine*, **8**, 543-576, 1996.
10. Pawlak, Z., *Rough Sets*. Kluwer Academic Publishers, Dordrecht, 1991.
11. Skowron, A. and Grzymala-Busse, J. From rough set theory to evidence theory. In: Yager, R., Fedrizzi, M. and Kacprzyk, J.(eds.) *Advances in the Dempster-Shafer Theory of Evidence*, pp.193-236, John Wiley & Sons, New York, 1994.
12. Tsumoto, S. and Tanaka, H., PRIMEROSE: Probabilistic Rule Induction Method based on Rough Sets and Resampling Methods. *Computational Intelligence*, **11**, 389-405, 1995.
13. Tsumoto, S. Knowledge Discovery in Medical MultiDatabases: A Rough Set Approach, *Proceedings of PKDD99*(in this issue), 1999.
14. Tsumoto, S. Mining Positive and Negative Knowledge in Clinical Databases based on Rough Set Model. *Proceedings of PKDD2001*.
15. Van Bommel, J. and Musen, M. A. *Handbook of Medical Informatics*, Springer-Verlag, New York, 1997.