Extraction of Sectorial Episodes Representing Changes for Drug Resistance and Replacements of Bacteria

Takashi Katoh
Kouichi Hirata
and Masateru Harao
Kyushu Institute of Technology
Iizuka 820-8502, Japan
Email: {f673024t,hirata,harao}@ai.kyutech.ac.jp

Shigeki Yokoyama
Koden Industrial Co., Ltd
Tokyo 143-0015, Japan
Email: shigey@st.rim.or.jp

Kimiko Matsuoka
Osaka Prefectural General Medical Center
Osaka 558-8558, Japan
Email: cby41060@pop01.odn.ne.jp

***Abstract*—A *sectorial episode* is of the form $C \mapsto r$, where C is a set of events and r is an event. Katoh *et al.* (2006) have designed the algorithm SECT to extract all of the sectorial episodes that are *frequent* and *accurate*. In this paper, by applying the algorithm SECT to bacterial culture data, we extract sectorial episodes representing *changes for drug resistance* and *replacements of bacteria*.**

I. INTRODUCTION

In Complex Medical Engineering, *data mining* is one of the key techniques to analyze medical data in order to extract medical information deeply. There exist many researches of data mining in medicine, that is, one from medical viewpoint, for example, [10].

In our previous works, we have paid our attention to *data mining from bacterial culture data*. Then, we have developed to extract *disjunctive rules*, called *frequent monotone DNF formulas* [3], *frequent closed monotone DNF formulas* [7] and *frequent few-overlapped monotone DNF formulas* [8], as hypotheses to explain bacterial culture data nearly overall. Also we have designed the algorithm to construct decision trees classifying MRSA to MSSA [2] and then implemented the prototype of risk management system for hospital-acquired infection [4].

However, since we have dealt with bacterial culture data like as itemsets [1], it remains open how to deal with bacterial culture data like as time-related data and how to extract frequent patterns representing temporal rules.

The *episode mining*, introduced by Mannila *et al.* [6], is known as one of the methods to discover frequent patterns from time-related data. The purpose of episode mining is to discover *frequent episodes* that are a collection of events occurring frequently together in event sequences.

In episode mining, the frequency is formulated as the number of occurrences of episodes in every *window* that is a subsequence of event sequences under a fixed time span called the *width* of windows. Then, Mannila *et al.* [6] have introduced a *parallel episode* as a set of events and a *serial episode* as a sequence of events. By combining the above episodes, they have extended the forms of episodes as *directed acyclic graphs* of events of which edges specify the temporal precedent-subsequent relationship.

As the form of an episode, Katoh *et al.* [5] have introduced a *sectorial episode* of the form $C \mapsto r$, where C is a set of event types and r is an event type. The sectorial episode $C \mapsto r$ means that every event of C is followed by an event r, so we can regard every event in C as a candidate of *causation* of r.

Then, they have designed the algorithm SECT [5] to extract *sectorial episodes that are frequent and accurate*. The main idea of the algorithm is to collect all of the frequent sectorial episodes of the form $C \mapsto r$ such that $|C| = 1$ and then store the information of windows in which $C \mapsto r$ occurs and then to construct all of the sectorial episodes based on the *antimonotonicity* of sectorial episodes.

Hence, in this paper, we regard bacterial culture data as event sequences, one of the time-related data. Then, we apply the algorithm SECT [5] to bacterial culture data. By the form $C \mapsto r$ of sectorial episodes, we can regard an element of C as cause of an event r.

The purpose of this paper is to extract the sectorial episodes representing *changes for drug resistance* and *replacements of bacteria*. In the former, we extract sectorial episodes of the form $C \mapsto r$ such that C contains an event denoting some antibiotics is susceptibility and an event r is an event denoting the antibiotics is resistant. In the latter, we extract sectorial episodes of the form $C \mapsto r$ such that C contains an event denoting the occurrence of bacterium and an event r is an event denoting the occurrence of different bacterium.

II. SECTORIAL EPISODE

Let $\mathcal{E}$ be a set of *event types*. Then, a pair (e, t) is called an *event*, where $e \in \mathcal{E}$ and t is a natural number which is the *occurrence time* of e. An *event sequence* on $\mathcal{E}$ is the set of events, where we deal with bacterial culture data as an event sequence.

A set of event types is a *parallel episode* [6]. Furthermore, let C be a parallel episode and r an event type. Then, a *sectorial episode* is of the form $C \mapsto r$.

We say that a parallel episode C *occurs* in $\mathcal{S}$ if every element of C occurs in $\mathcal{S}$. Also we say that a sectorial episode $C \mapsto r$ *occurs* in $\mathcal{S}$ if there exists an r that occurs in $\mathcal{S}$ such that every element of C occurs in $\mathcal{S}$ and is precedent to this r.

For an event sequence $\mathcal{S}$, a subset $\{(e,t) \in \mathcal{S} \mid t_s \leq t < t_e\}$ of $\mathcal{S}$ is called a *window* in $\mathcal{S}$ with *width* $t_e - t_s$. We denote the set of all windows in $\mathcal{S}$ with width k by $W(k)$, and the set of all windows in $\mathcal{S}$ with width k in which an episode X occurs by $W(k,X)$.

Let $\mathcal{S}$ be an event sequence k an integer. By regarding a sectorial episode $X = C \mapsto r$ as an association rule [1], we can formulate the *support* $supp_{\mathcal{S},k}(X)$ and the *confidence* $conf_{\mathcal{S},k}(X)$ of X as follows:

$$supp_{\mathcal{S},k}(X) = \frac{|W(k,X)|}{|W(k)|}, \quad conf_{\mathcal{S},k}(X) = \frac{|W(k,X)|}{|W(k,C)|}.$$

In the following, the subscripts $\mathcal{S}$ and k are omitted if they are clear by the context.

For the *minimum support* σ $(0 < \sigma < 1)$, we say that a sectorial episode X is *frequent* if $supp(X) \geq \sigma$. Also, for the *minimum confidence* γ $(0 < \gamma < 1)$, we say that a sectorial episode X is *accurate* if $conf(X) \geq \gamma$.

A sectorial episode is *anti-monotonic*, that is:

> *For parallel episodes C_1 and C_2 such that $C_1 \subseteq C_2$, if $C_2 \mapsto r$ is frequent, then so is $C_1 \mapsto r$.*

Based on the antimonotonicity of sectorial episodes, Katoh *et al.* [5] have designed the algorithm SECT to extract all of the sectorial episodes that are frequent and accurate. The oultine of the algorithm SECT is described as follows.

1) The algorithm SECT first collects the information of windows in which $C \mapsto r$ such that $|C| = 1$ occurs, by traversing a give event sequence just once.
2) Then, it constructs sectorial episodes from them in breadth-first search with pruning infrequent ones based on the anti-monotonicity.

III. EXPERIMENTAL RESULTS

In this section, we apply the algorithm SECT to bacterial culture data from Osaka Prefectural General Medical Center from 1995 to 1998, which are complete data in [9]. Then, we extract sectorial episodes representing *changes for drug resistance* and *replacements of bacteria*. Here, we regard a pair of "attribute=value" as an event type.

In the following, we describe the event sequence by using the parameters (d,p,w), where d is the number of records, p is the number of patients and w is the number of windows, respectively. Furthermore, we denote the number of extracted sectorial episodes by e.

A. Changes for drug resistance

First, we extract sectorial episodes representing changes for drug resistance[1]. By fixing the category of bacteria and the sample, we connect data of every patient with the span of 30 days, which is the width of windows. Then, we investigate the sectorial episode of the following form:

$$C \mapsto (\texttt{Ant=R}),$$

where $(\texttt{Ant=S}) \in C$. The attribute values `S` and `R` denote "susceptibility" and "resistant," respectively. We call such an episode a *target episode* (of an antibiotic `Ant`), which means that the drug resistance of `Ant` changes from `S` to `R`.

In this section, $\texttt{Ant}(n)$ denotes that the number of extracted target episodes of an antibiotic `Ant` is n. Here, antibiotics are benzilpenicillin (`PcB`), augmentin (`Aug`), anti-pseudomonas penicillin (`PcAP`), 1st generation cephems (`Cep1`), 2nd generation cephems (`Cep2`), 3rd generation cephems (`Cep3`), aminoglycosides (`AG`), macrolides (`ML`), tetracyclines (`TC`), and carbapenems (`CBP`).

1) Staphylococci: Consider the case that the category of bacteria is "Staphylococci." Then, we can extract episodes from the samples "catheter/others," "respiratory organs," and "blood." The following table describes the parameters of the event sequence for the above samples.

sample	d	p	w
catheter/others	775	249	41521
respiratory organs	1014	311	46062
blood	89	33	2830

For the first two samples, we set the minimum support $\sigma = 0.01$. Then, for the minimum confidence γ, we obtain the following target episodes.

sample	γ	e	antibiotics
catheter/others	0.05	24	`PcS`(16), `Cep1`(8)
	0.10	0	–
respiratory organs	0.05	10776	`ML`(4672), `Cep1`(3792), `PcS`(2128), `TC`(184)
	0.10	5644	`Cep1`(3512), `PcS`(1860), `ML`(272)

On the other hand, for the sample "blood," we set the minimum support $\sigma = 0.04$ and 0.03. Then, for the minimum confidence γ, we obtain the following target episodes.

σ	γ	e	antibiotics
0.04	0.10	560	`AG`(544), `CBP`(16)
	0.15	16	`AG`(16)
0.03	0.10	7208	`AG`(3008), `PcS`(2048), `Cep1`(1728), `CBP`(400), `TC`(24)
	0.15	4992	`PcS`(2048), `Cep1`(1728), `AG`(976), `CBP`(224), `TC`(16)

Hence, for the above three samples, we can extract frequent target episodes of `PcS` and `Cep1` in "Staphylococci." Also, for the sample "blood," we can extract ones of `AG` and `CBP`. On the other hand, the frequency of target episodes of `ML` for

[1] In the previous work [5], we have already reported the experimental results from the same data, but the implementation of the algorithm SECT contains some errors. Hence, the result given in this section is different from one given by [5].

the sample "respiratory organs" and AG for the sample "blood" rapidly decreases when the minimum confidence increases.

2) Enterococci: Consider the case that the category of bacteria is "Enterococci." Then, we can extract episodes from the samples "catheter/others" and "urinary/genital organs." The following table describes the parameters the event sequence for the above samples.

sample	d	p	w
catheter/others	206	70	9318
urinary/genital organs	120	45	7601

We set the minimum support $\sigma = 0.01$. Then, for the minimum confidence γ, we obtain the following target episodes.

sample	γ	e	antibiotics
catheter/others	0.05	2112	PcB(1792), TC(320)
	0.10	640	PcB(512), TC(128)
urinary/genital organs	0.05	96	PcB(96)
	0.10	0	–

Hence, for the above two samples, we can extract frequent target episodes of PcB and TC in "Enterococci," where the former is more frequent than the latter.

3) Enteric bacteria: Consider the case that the category of bacteria is "Enteric bacteria." Then, we can extract episodes from the samples "catheter/others" and "respiratory organs." The following table describes the parameters of the event sequence for the above samples.

sample	d	p	w
catheter/others	441	133	14379
respiratory organs	987	308	39917

For the sample "catheter/others," we set the minimum support $\sigma = 0.04$ and 0.03. Then, for the minimum confidence γ, we obtain the following target episodes.

σ	γ	e	antibiotics
0.04	0.10	384	Aug(384)
	0.15	352	Aug(352)
0.03	0.10	38784	Aug(32768), Cep1(6016)
	0.15	9764	Aug(9764)

On the other hand, for the sample "respiratory organs," we set the minimum support $\sigma = 0.03$ and 0.02. Then, for the minimum confidence γ, we obtain the following target episodes.

σ	γ	e	antibiotics
0.02	0.10	168396	Aug(78832), Cep1(76016), Cep2(13548)
	0.15	42104	Aug(23152), Cep1(18952)
0.03	0.10	1536	Aug(880), Cep1(656)
	0.15	0	–

Hence, for the above two samples, we can extract frequent target episodes of Aug and Cep1 in "Enteric bacteria," where the former is more frequent than the latter.

4) Glucose-nonfermentative gram-negative bacteria: Consider the case that the category of bacteria is "Glucose-nonfermentative gram-negative bacteria." Then, we can extract episodes from the samples "catheter/others" and "respiratory organs." The following table describes the parameters of the event sequence for the above samples.

sample	d	p	w
catheter/others	238	81	12585
respiratory organs	1030	302	48195

For these samples, we set the minimum support $\sigma = 0.01$. Then, for the minimum confidence γ, we obtain the following target episodes.

sample	γ	e	antibiotics
catheter/others	0.05	1096	CBP(1088), TC(8)
	0.10	372	CBP(368), TC(8)
	0.15	0	–
respiratory organs	0.05	15176	CBP(12224), TC(2088), Cep3(544), Aug(320)
	0.10	14260	CBP(11788), TC(2088), Aug(320), Cep3(64)
	0.15	6652	CBP(4308), TC(2024), Aug(320)

Hence, for the above two samples, we can extract frequent target episodes of CBP and TC in "Glucose-nonfermentative gram-negative bacteria," where the former is more frequent than the latter. Also, for the sample "respiratory organs," we can extract ones of Aug, but they are more infrequent than ones of CBP and TC.

5) Other gram-negative bacilli: Consider the case that the category of bacteria is "Other gram-negative bacilli." Then, we can extract episodes from the sample "respiratory organs." The parameters of the event sequence for the sample "respiratory organs" is that $(d, p, w) = (484, 179, 32759)$.

We set the minimum support $\sigma = 0.01$. Then, for the minimum confidence $\gamma = 0.05$ and 0.10, we obtain the target episodes of which antibiotics is PcB(6524) and PcB(1536), respectively. Also for the minimum confidence $\gamma = 0.15$, we can obtain no target episodes.

Hence, for the sample "respiratory organs," we can extract frequent target episodes of just PcB in "Other gram-negative bacilli."

6) Anaerobes: Consider the case that the category of bacteria is "Anaerobes." Then, we can extract episodes from the samples "catheter/others," "punctual fluid," and "blood." The following table describes the parameters of the event sequence for the above samples.

sample	d	p	w
catheter/others	763	218	19032
punctual fluid	70	27	941
blood	114	45	1785

For the sample "catheter/others," we set the minimum support $\sigma = 0.05$ and 0.04. Then, for the minimum confidence γ, we obtain the following target episodes.

σ	γ	e	antibiotics
0.05	0.10	416	Cep1(352), PcB(64)
	0.15	192	Cep1(128), PcB(64)
0.04	0.10	41152	PcB(27008), Cep1(14144)
	0.15	14216	PcB(13408), Cep1(808)

For the last two samples "punctual fluid" and "blood," we set the minimum support $\sigma = 0.05$. Then, for the minimum confidence γ, we obtain the following target episodes.

sample	γ	e	antibiotics
punctual fluid	0.10	114688	PcB(57344), Cep1(57344)
	0.30	65536	PcB(32768), Cep1(32768)
	0.50	16384	PcB(8192), Cep1(8192)
blood	0.05	1728	PcAP(1216), ML(512)
	0.10	1088	PcAP(576), ML(512)
	0.15	512	PcAP(512)

Hence, for the samples "catheter/others" and "punctual fluid," we can extract frequent target episodes of PcB and Cep1 in "Anaerobes." On the other hand, for the sample "blood," we can extract frequent target episodes of PcAP and ML in "Anaerobes."

Note that, for the sample "punctual fluid," the minimum confidence is very large than other experimental results in this section. Furthermore, in this case, the frequency of target episodes of PcB is same as ones of Cep1 for every minimum confidence.

7) Summary: Finally, we summarize the above experimental results for the viewpoint of samples. The following table describes the changes for drug resistance for every samples.

sample	bacterium	antibiotics
catheter/others	Staphylococci	PcS, Cep1
	Enterococci	PcB, TC
	Enteric bacteria	Aug, Cep1
	Glucose-nonfermentative gram-negative bacteria	TC, CBP
	Anaerobes	PcB, Cep1
punctual fluid	Anaerobes	PcB, Cep1
urinary/genital organs	Enterococci	PcB
respiratory organs	Staphylococci	Cep1, ML PcS, TC
	Enteric bacteria	Aug, Cep1, Cep2
	Glucose-nonfermentative gram-negative bacteria	Aug, Cep3, TC, CBP
	other gram-negative bacilli	PcB
blood	Staphylococci	PcS, Cep1, AG, TC, CBP
	Anaerobes	PcAP, ML

The previous work [11] has reported that the drug resistance of CBP changes from S to R in Anaerobes from this data. While we can observe no changes for drug resistance of CBP in "Anaerobes," we can newly observe ones in "Glucose-nonfermentative gram-negative bacteria" for the samples "catheter/others" and "respiratory organs," and in "Staphylococci" for the sample "blood." Furthermore, in "Anaerobes," we can newly observe the changes for the drug resistance of PcB and Cep1 for the samples "catheter/others" and "punctual fluid," and ones of PcAP and ML for the sample "blood."

B. Replacements of bacteria

Next, we extract the sectorial episodes representing *replacements of bacteria*. By fixing the sample, we connect data of every patient with the span of 30 days. Then, the following table describes the parameters of the event sequence for samples.

sample	d	p	w
catheter/others	3642	834	94568
punctual fluid	226	67	3322
urinary/genital organs	1152	386	39139
digestive tract	418	141	11465
respiratory organs	5639	1337	170604
blood	468	162	8986

In this section, we pay our attention to the sectorial episode of the following form:

$$C \mapsto (\texttt{Bac=bac2}),$$

where $(\texttt{Bac=bac1}) \in C$ such that $\texttt{bac1} \neq \texttt{bac2}$. We call the attribute values bac1 and bac2 a *precedent bacterium* and a *subsequent bacterium*, respectively. We call such an episode a *target episode*.

1) Catheter/others: Consider the case that the sample is "catheter/others." We set the minimum support $\sigma = 0.005$ and the minimum confidence $\gamma = 0.10$. Then, we obtain the following 215248 target episodes.

precedent	subsequent	e
Enterococcus faecalis	Pseudomonas aeruginosa	214768
Bacteroides fragilis	Enterococcus faecalis	480

Note that we can obtain no target episodes under the minimum confidence $\gamma = 0.20$.

2) Punctual fluid: Consider the case that the sample is "punctual fluid." We set the minimum support $\sigma = 0.0175$. In this case, we just extract the following target episodes.

precedent	subsequent
Enterococcus faecalis	Bacteroides fragilis

For the minimum confidence γ, the number of target episodes is 65536 ($\gamma = 0.10$), 56832 ($\gamma = 0.20$), 18432 ($\gamma = 0.30$) and 0 ($\gamma = 0.40$).

3) Urinary/genital organs: Consider the case that the sample is "urinary/genital organs." We set the minimum support $\sigma = 0.0025$. Then, we obtain the following target episodes.

$\gamma = 0.10$

precedent	subsequent	e
Klebsiella pneumoniae	Enterococcus faecalis	24448
Morganella morganii	Pseudomonas aeruginosa	16352
Citrobacter freundii	Enterococcus faecalis	4072
Staphylococcus aureus	Enterococcus faecalis	3352
Enterococcus faecalis	Pseudomonas aeruginosa	1552
Pseudomonas aeruginosa	Enterococcus faecalis	384

$\gamma = 0.20$

precedent	subsequent	e
Morganella morganii	Pseudomonas aeruginosa	8192
Staphylococcus aureus	Enterococcus faecalis	256

$\gamma = 0.30$

precedent	subsequent	e
Morganella morganii	Pseudomonas aeruginosa	8192

For the minimum confidence γ, the total number of target episodes is 50160 ($\gamma = 0.10$), 8448 ($\gamma = 0.20$), 8192 ($\gamma = 0.30$) and 0 ($\gamma = 0.40$).

4) Digestive tract: Consider the case that the sample is "digestive tract." We set the minimum support $\sigma = 0.0075$. In this case, we just extract the following target episodes.

precedent	subsequent
Staphylococcus aureus	Pseudomonas aeruginosa

For the minimum confidence γ, the number of target episodes is 11168 ($\gamma = 0.10$), 3016 ($\gamma = 0.20$), 1272 ($\gamma = 0.30$), 736 ($\gamma = 0.40$), 240 ($\gamma = 0.50$) and 0 ($\gamma = 0.60$).

5) Respiratory organs: Consider the case that the sample is "respiratory organs." We set the minimum support $\sigma = 0.01$. Then, we obtain the following target episodes.

$\gamma = 0.10$

precedent	subsequent	e
Pseudomonas aeruginosa	Staphylococcus aureus	517480
Staphylococcus aureus	Pseudomonas aeruginosa	149100

$\gamma = 0.20$

precedent	subsequent	e
Pseudomonas aeruginosa	Staphylococcus aureus	71712

For the minimum confidence γ, the number of target episodes is 666580 ($\gamma = 0.10$), 71712 ($\gamma = 0.20$) and 0 ($\gamma = 0.30$).

6) Blood: Consider the case that the sample is "blood." We set the minimum support $\sigma = 0.02$. In this case, we just extract the following target episodes.

precedent	subsequent
Streptococcus intermedius	Staphylococcus aureus

For the minimum confidence γ, the number of target episodes is 16384 ($\gamma = 0.10$), 16384 ($\gamma = 0.20$), 8128 ($\gamma = 0.30$) and 0 ($\gamma = 0.40$).

7) Summary: We summarize the above experimental results. The following table describes the replacements of bacterium for every samples.

sample	precedent	subsequent
catheter/others	Bacteroides fragilis	Enterococcus faecalis
	Enterococcus faecalis	Pseudomonas aeruginosa
punctual fluid	Enterococcus faecalis	Bacteroides fragilis
urinary/genital organs	Enterococcus faecalis	Pseudomonas aeruginosa
	Morganella morganii	Pseudomonas aeruginosa
	Pseudomonas aeruginosa	Enterococcus faecalis
	Staphylococcus aureus	Enterococcus faecalis
	Citrobacter freundii	Enterococcus faecalis
	Klebsiella pneumoniae	Enterococcus faecalis
digestive tract	Staphylococcus aureus	Pseudomonas aeruginosa
respiratory organs	Pseudomonas aeruginosa	Staphylococcus aureus
	Staphylococcus aureus	Pseudomonas aeruginosa
blood	Streptococcus intermedius	Staphylococcus aureus

Hence, the bacteria "Enterococcus faecalis," "Staphylococcus aureus," "Pseudomonas aeruginosa," and "Bacteroides fragilis" occur in a target episode as a precedent or a subsequent bacteria.

IV. Conclusion

In this paper, by applying the algorithm SECT [5] to bacterial culture data, we have extracted the sectorial episodes, that are frequent and accurate, representing changes for drug resistance and replacements of bacteria.

Since the number of extracted sectorial episodes in Section III is large, it is a future work to introduce the concept of *closed* sectorial episodes, in order to reduce the number of extracted episodes.

Furthermore, the number of extracted sectorial episodes tends to be large when the number of windows is small. Hence, it is a future work to give the experimental results by introducing the method to align the number of windows.

In Section III, we have connected data of every patient with the span of 30 days and adopted the frequency measure based on the number of windows. However, it is possible to count the similar windows repeatedly, so it is a future work to introduce another frequency measure like as the frequency for every patient.

Finally, it is an important future work to analyze the experimental results of this paper deeply from the medical viewpoints and to investigate the relationship between the results and the hospital infection.

References

[1] R. Agrawal, R. Srikant: *Fast algorithms for mining association rules in large databases*, Proc. 20th VLDB, 487–499, 1994.

[2] K. Hirata, M. Harao, M. Wada, S. Ozaki, S. Yokoyama, K. Matsuoka: *Attribute selection measures with possibility and their application to classifying MRSA from MSSA*, in [10], 143–151.

[3] K. Hirata, Y. Shima, M. Harao, S. Yokoyama, K. Matsuoka, T. Izumi: *Disjunctive rules extracted from MRSA data with verification*, Proc. CME'05, 326–330, 2005.

[4] Y. Ikenaga, K. Hirata, M. Harao, S. Yokoyama, K. Matusoka: *Risk management system for hospital-acquired infection based on bacterial culture database*, Proc. CME'05, 423–427, 2005.

[5] T. Katoh, K. Hirata, M. Harao: *Mining sectorial episodes from event sequences*, Proc. DS'06, LNAI **4265**, 137–145, 2006.

[6] H. Mannila, H. Toivonen, A. I. Verkamo: *Discovery of frequent episodes in event sequences*, Data Mining and Knowledge Discovery **1**, 259–289, 1997.

[7] Y. Shima, K. Hirata, M. Harao, S. Yokoyama, K. Matsuoka, T. Izumi: *Extracting disjunctive closed rules from MRSA data*, Proc. CME'05, 321–325, 2005.

[8] Y. Shima, K. Hirata, M. Harao: *Extraction of frequent few-overlapped monotone DNF formulas with depth-first prunning*, Proc. PAKDD'05, LNAI **3518**, 50–60, 2005.

[9] S. Tsumoto: *Guide to the bacteriological examination data set*, in E. Suzuki (ed.): Proc. International Workshop of KDD Challenge on Real-World Data (KDD Challenge 2000), 8–12, 2000.

[10] J. L. Wu, K. Ito, S. Tobimatsu, T. Nishida, H. Fukuyama (eds.): *Complex medical engineering*, Springer, 2007.

[11] S. Yokoyama, K. Matsuoka, S. Tsumoto, M. Harao, T. Yamakawa, K. Sugahara, C. Nakahama, S. Ichiyama, K.Watanabe: *Study on the association between the patients' clinical background and the Anaerobes by data mining in infectious diseases database*, Biomedical Soft Computing and Human Sciences **7**, 69–75, 2001.