

Analysis of Hepatitis Dataset by Decision Tree Graph-Based Induction ^{*}

Kouzou Ohara^{*}, Tetsuya Yoshida^{**},
Warodom Geamsakul^{*}, Hiroshi Motoda^{*}, Takashi Washio^{*},
Hideto Yokoi^{***}, and Katsuhiko Takabayashi^{***}

^{*} Institute of Scientific and Industrial Research, Osaka University, JAPAN
{ohara,warodom,motoda,washio}@ar.sanken.osaka-u.ac.jp

^{**} Graduate School of Information Science and Technology,
Hokkaido University, JAPAN
yoshida@meme.hokudai.ac.jp

^{***} Division for Medical Informatics, Chiba University Hospital, JAPAN
yokoi@telemed.ho.chiba-u.ac.jp,takaba@ho.chiba-u.ac.jp

Abstract. We analyzed the hepatitis data by Decision Tree Graph-Based Induction (DT-GBI), which constructs a decision tree for graph-structured data while simultaneously constructing attributes for classification. An attribute at each node in the decision tree is a discriminative pattern (subgraph) in the input graph, and extracted by Graph-Based Induction (GBI). We conducted four kinds of experiments using only the time-series data of blood inspection and urinalysis. In the first and second experiments, the stages of fibrosis were used as classes and a decision tree was constructed for discriminating patients with F4 (cirrhosis) from patients with the other stages. In the third experiment, the types of hepatitis (B and C) were used as classes, and in the fourth experiment the effectiveness of interferon therapy was used as the class label. The preliminary results of experiments, both constructed decision trees and their predictive accuracies, are reported in this paper.

1 Introduction

Chronic Hepatitis data which was provided by Chiba University Hospital contains irregular and multi-dimensional long time-series data. Due to its comprehensibility, it is very difficult for medical experts to analyze the data and discover useful medical knowledge by themselves, and systematic methods which can support them are desired.

There are some systematic analyses already conducted and reported on this dataset. [9] analyzed the data by constructing decision trees from time-series data without discretizing numeric values. [1] proposed a method of temporal abstraction to handle time-series data, converted time phenomena to symbols and used a standard classifier. [4] used multi-scale matching to compare time-series

^{*} The results in this paper have been included in the working notes of the Second and Third International Workshops on Active Mining(AM2003 and AM2004)[7, 11].

data and clustered them using rough set theory. [3] also clustered the time-series data of a certain time interval into several categories and used a standard classifier. These analyses examine the temporal correlation of each inspection separately and do not explicitly consider the relations among inspections. Thus, these approaches do not correspond to the structured data analysis. As another approach, [8] visualized the hepatitis data as a sequence of probabilistic prototypes. This method allows even a non-expert in medicine to discover interesting tendencies from the comprehensible data, although it does not generate any hypotheses by itself.

We analyzed the hepatitis data by Decision Tree Graph-Based Induction (DT-GBI)[5, 6], which constructs a classifier (decision tree) for graph-structured data while simultaneously constructing attributes for classification using Graph-based Induction (GBI) [10, 2]. A pair extracted by GBI, which consists of nodes and the edges between them, is treated as an attribute and the existence/non-existence of the pair in a graph is treated as its value for the graph. Attributes (pairs) that divide data effectively are extracted by GBI, while a decision tree is being constructed.

In our analysis temporal records in the provided dataset was converted into graph-structured data with respect to time correlation so that both intra-correlation of an individual inspection at one time and inter-correlation among inspections at different time are represented in graph-structured data. The decision trees were constructed by DT-GBI to discriminate between two groups of patients without using any biopsy results but using only the time sequence data of blood inspection and urinalysis. The stages of fibrosis were used as classes in the first and second experiments to discriminate patients with F4 (cirrhosis) from patients with the other stages. The types of hepatitis (B and C) were used as classes in the third experiment, and the effectiveness of interferon therapy was used as class label in the fourth experiment.

Section 2 briefly describes the framework of GBI. Section 3 explains the preprocessing of the data. Section 4 reports the results of experiments, both constructed decision trees and their predictive accuracies. Section 5 concludes the paper with a summary of the results and the planned future work.

2 Decision Tree by GBI (DT-GBI)

Decision Tree Graph-based Induction (DT-GBI) constructs a decision tree for graph-structured data while simultaneously constructing attributes for classification using Graph-Based Induction (GBI), which employs the idea of extracting typical patterns by stepwise pair expansion (we call this process “chunking”). Chunking integrates a pair of nodes into a new node, which represents a pattern consisting of them. In GBI, assumptions are made that typical patterns represent some concepts and “typicality” is characterized by the pattern’s frequency or the value of some evaluation function based on its frequency. Repeated chunking enables GBI to extract typical or discriminative patterns of various sizes.

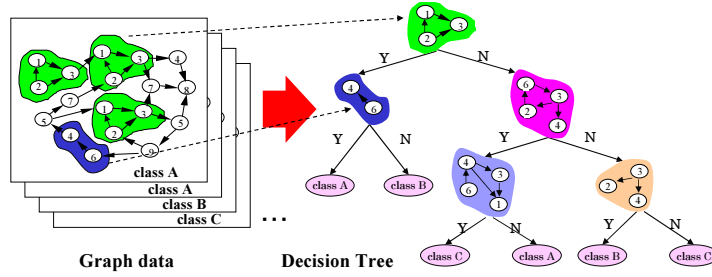


Fig. 1. Decision tree for classifying graph-structured data

DT-GBI employs discriminative patterns (subgraphs) as attributes in a decision tree, and recursively extracts them from the input graphs by GBI. Namely we can represent graph-structured data with attribute value pairs according to the existence of a particular subgraph: the value for an attribute (subgraph) of a graph is *yes* if the graph contains the subgraph; otherwise *no*. The resulting decision tree becomes a binary tree. Chunking is applied for a specified number of times at each node of a decision tree and the chunked pairs grow up into larger nodes in size. Thus, although initial pairs consist of only two nodes and one edge between them, attributes useful for classification task are gradually grown up into larger pairs (subgraphs) by applying chunking recursively. The above process is illustrated in Fig.1. To classify graph-structured data after the construction of a decision tree, attributes are produced from data before the classification.

Note that the search in GBI is greedy and no backtracking is made because its objective is not to find all typical patterns nor to find all frequent patterns, but to extract only meaningful typical patterns of certain sizes. Consequently there can be many patterns which are not extracted by GBI. Thus DT-GBI actually adopts an extension of GBI called Beam-wise Graph-Based Induction (B-GBI)[2] to construct discriminative patterns. B-GBI increases the search space by introducing beam search, and can extract more discriminative patterns while keeping the computational complexity within a tolerant level. Given a beam width b , B-GBI selects the best b pairs to be chunked individually in parallel. As for the details of GBI, B-GBI, and DT-GBI, please refer to [10, 2, 5, 6].

3 Data Preprocessing

The hepatitis data contains long time-series data (from 1982 to 2001) on laboratory examination of 771 patients of hepatitis B and C, and can be split into two categories. The first category includes administrative information such as patient’s information (age and date of birth), pathological classification of the disease, date of biopsy and its result, and the effectiveness of interferon therapy. The second category includes temporal record of blood test and urinalysis. It contains the result of 983 types of both in- and out-hospital examinations.

MID	Date of examination	Object to examine	Name of examination	...	Result value	Unit	Judge result	Comment	...
1	19850711	1	CA19-9		8	U/ML			...
1	19870114	1	CMVJGG(ELISA)		0.729		(2+)		...
1	19870114	1	CMVJGM(ELISA)		0.214		(-)	サイケンスミデス	...
...
2	19920611	1	2-5ASカンセイ		69	PMOL/DL			...
2	19941003	1	HCVSNGR RT-PCR		(3+)				...
2	19950911	1	HCVタイプヨウ(プローブ)		6.5	MEQ/ML			...
...

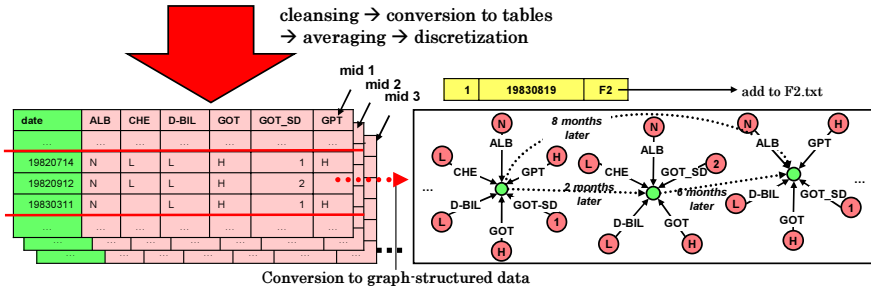


Fig. 2. Outline of the preprocessing of inspection data

Cleansing and Conversion to Tables In numeric attributes, letters and symbols are deleted. Values in nominal attributes are left as they are. When converting the given data into attribute-value tables for each patient, both a patient ID (MID) and a date of inspection are used as search keys and an inspection item is defined as an attribute. Since it is not necessary that all patients must take all inspections, there are many missing values after this data conversion. No attempt is made to estimate these values and those missing values are not represented in graph-structured data in the following experiments.

Averaging and Discretization This step is necessary due to the following two reasons: 1) obvious change in inspection across successive visits may not be found because the progress of hepatitis is slow, and 2) the date of visit is not synchronized across different patients. In this step, the numeric attributes are averaged and the most frequent value is used for nominal attributes over some interval. Further, for some inspections (GOT, GPT, TTT, and ZTT), standard deviations are calculated over six-month and added as new attributes.

In graph-structured data, we represent an inspection result as a node label. To reduce the number of node labels, we discretized attribute values because the large number of node labels could prevent DT-GBI from properly extracting frequent and discriminative patterns. For general numerical values, the normal ranges are specified and values are discretized into three (“L” for low, “N” for normal, and “H” for high). Based on the range, the standard deviations of GOT and GPT are discretized into five (“1” for the smallest deviation, “2”, “3”, “4”, “5” for the largest deviation), while the standard de-

viations of TTT and ZTT are discretized into three (“1” for the smallest deviation, “2”, “3” for the largest deviation). Figure 2 shows the outline of the data preprocessing, representing the four steps mentioned above, i.e., cleansing, conversion to tables, averaging, and discretization.

Limiting Data Range We assume that each patient has one class label, which is determined at some inspection date. The longer the interval between the date when the class label is determined and the date of blood inspection is, the less reliable the correlation between them is. We consider that the pathological conditions remains the same for some duration and conduct the analysis for the data which lies in the range. Furthermore, feature selection was conducted with the expert to reduce the number of attributes. Then the following 32 attributes were selected out of the 987 examinations: ALB, CHE, D-BIL, GOT, GOT_SD, GPT, GPT_SD, HBC-AB, HBE-AB, HBE-AG, HBS-AB, HBS-AG, HCT, HCV-AB, HCV-RNA, HGB, I-BIL, ICG-15, MCH, MCHC, MCV, PLT, PT, RBC, T-BIL, T-CHO, TP, TTT, TTT_SD, WBC, ZTT, and ZTT_SD. Indeed, attributes were further limited according to the objective of the analysis. The duration and attributes actually used in each analysis are described in the following results.

Conversion to Graph-Structured Data To analyze data by DT-GBI, it is necessary to convert the data to graph structure. One patient record is mapped into one directed graph. Figure 2 includes an example of converted graph-structured data. In this figure, a star-shaped subgraph represents values of a set of pathological examination in a two-month interval. The center node of the subgraph is a hypothetical node for the interval. An edge pointing to a hypothetical node represents an examination. The node connected to the edge represents the value (preprocessed result) of the examination. The edge linking two hypothetical nodes represents time difference. We considered direct time correlation between two sets of pathological tests only within a predefined interval (here, two years).

Class Label Setting In the first and second experiments, we set the result (progress of fibrosis) of the first biopsy as class. In the third experiment, we set the subtype (B or C) as class. In the fourth experiment, the effectiveness of interferon therapy was used as class label.

4 Preliminary Results

4.1 Initial Settings

A decision tree was constructed in either of the following two ways: 1) apply chunking $n_r=20$ times at the root node and only once at the other nodes of a decision tree, 2) apply chunking $n_e=20$ times at every node of a decision tree. Decision tree pruning is conducted by postpruning: conduct pessimistic pruning by setting the confidence level to 25%. We evaluated the prediction accuracy of decision trees by the average of 10 runs of 10-fold cross-validation. Thus, 100 decision trees were constructed in total. In each experiment, to determine an optimal beam width b , we first conducted 9-fold cross-validation for one randomly

Table 1. Size of graphs (fibrosis stage)

stage of fibrosis	F0	F1	F2	F3	F4	Total
No. of graphs	4	125	53	37	43	262
Avg. No. of nodes	303	304	308	293	300	303
Max. No. of nodes	349	441	420	414	429	441
Min. No. of nodes	254	152	184	182	162	152

Table 2. Average error rate (%) (fibrosis stage)

run of 10 CV	F4 vs. {F0,F1}		F4 vs. {F2,F3}	
	$n_r=20$	$n_e=20$	$n_r=20$	$n_e=20$
1	14.81	11.11	27.78	25.00
2	13.89	11.11	26.85	25.93
3	15.74	12.03	25.00	19.44
4	16.67	15.74	27.78	26.68
5	16.67	12.96	25.00	22.22
6	15.74	14.81	23.15	21.30
7	12.96	9.26	29.63	25.93
8	17.59	15.74	25.93	22.22
9	12.96	11.11	27.78	21.30
10	12.96	11.1	27.78	25.00
average	15.00	12.50	26.67	23.52
Standard Deviation	1.65	2.12	1.80	2.39

chosen 9 folds data (90% of all data) of the first run of 10-fold cross-validation varying b from 1 to 15, and adopted the narrowest beam width that brings to the lowest error rate.

In the following subsections, both the average error rate and examples of decision trees are shown in each experiment. Two decision trees were selected as examples out of the 100 decision trees in each experiment: one from the 10 trees constructed in the *best run* with the lowest error rate of the 10 runs of 10-fold cross validation, and the other from the 10 trees in the *worst run* with the highest error rate. In addition, we show the overall contingency table, which is calculated by summing up the contingency tables for the 100 decision trees in each experiment.

4.2 Experiments 1 and 2: Classifying Patients with Fibrosis Stages

For the analysis in subsection 4.2 and subsection 4.3, the average was taken for two-month interval. As for the duration of data considered, data in the range from 500 days before to 500 days after the first biopsy were extracted for the analysis of biopsy result. When biopsy was operated for several times on the same patient, the treatment (*e.g.*, interferon therapy) after a biopsy may influence the result of blood inspection and lower the reliability of data. Thus, the date of first biopsy and the result of each patient are searched from the biopsy data file. In case that the result of the second biopsy or after differs from the result of the first one, the result from the first biopsy is defined as the class of that patient for the entire 1,000-day time-series.

Fibrosis stages are categorized into five stages: F0 (normal), F1, F2, F3, and F4 (severe = cirrhosis). We constructed decision trees which distinguish the patients at F4 stage from the patients at the other stages. In the following two experiments, we used all the 32 attributes. Table 1 shows the size of graphs after the data conversion. As shown in Table 1, the number of instances (graphs) in cirrhosis (F4) stage is 43 while that in non-cirrhosis stages (F0 + F1 + F2 +

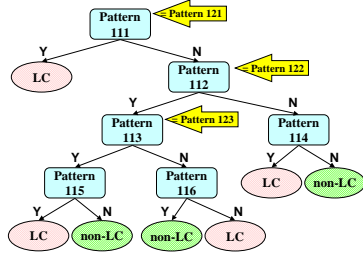


Fig. 3. One of the ten trees from the best run in exp.1 ($n_e=20$)

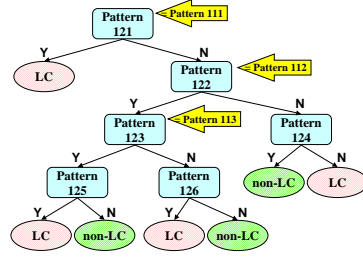


Fig. 4. One of the ten trees from the worst run in exp.1 ($n_e=20$)

Table 3. Contingency table with the number of instances (F4 vs. {F0+F1})

Actual Class	Predicted Class	
	LC	non-LC
LC	364	66
non-LC	69	581

LC = F4, non-LC = {F0+F1}

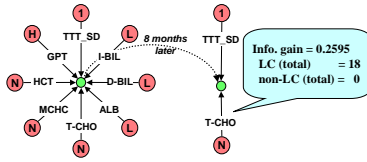


Fig. 5. Pattern 111 = Pattern 121

F3) is 219. Then we limited the number of instances to the 2:3 (cirrhosis:non-cirrhosis) ratio which is the same as in [9]. Thus, we used all instances from F4 stage for cirrhosis class (represented as LC) and select 65 instances from the other stages for non-cirrhosis class (represented as non-LC), 108 instances in all. How we selected these 108 instances is described below.

Experiment 1: F4 stage vs {F0+F1} stages All 4 instances in F0 and 61 instances in F1 stage were used for non-cirrhosis class, and b was set to 15 in this experiment. In this experiment, the overall result is summarized in the left half of Table 2. The average error rate was 15.00% for $n_r=20$ and 12.50% for $n_e=20$. Figures 3 and 4 show the decision trees constructed in the best and worst runs for $n_e=20$ (run 7 and run 8), respectively. Comparing these two decision trees, we notice that three patterns that appeared at the upper levels of each tree are identical. Figure 5 shows the pattern appeared at the root node of both trees. The overall contingency table for this experiment is shown in Table 3. Although the number of misclassified instances for LC (F4) and non-LC ({F0+F1}) are almost the same, since the class distribution of LC and non-LC is unbalanced (note that the number of instances is 65 for LC and 108 for non-LC), the error rate for LC was larger than that for non-LC.

Experiment 2: F4 stage vs {F3+F2} stages In this experiment, we used all instances in F3 and 28 instances in F2 stage for non-cirrhosis class, and b was set to 14. The overall result is summarized in the right-hand side of Table 2. The average error rate was 26.67% for $n_r=20$ and 23.52% for

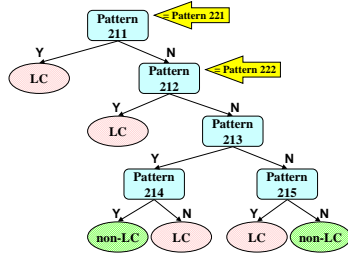


Fig. 6. One of the ten trees from the best run in exp.2 ($n_e=20$)

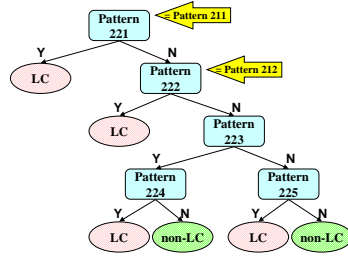


Fig. 7. One of the ten trees from the worst run in exp.2 ($n_e=20$)

Table 4. Contingency table with the number of instances (F4 vs. {F3+F2})

Actual Class	Predicted Class	
	LC	non-LC
LC	282	148
non-LC	106	544

LC = F4, non-LC = {F3+F2}

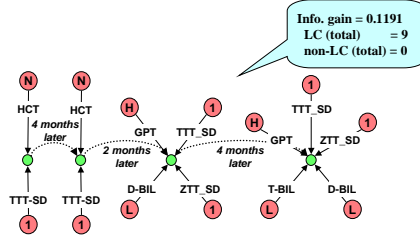


Fig. 8. Pattern 211 = Pattern 221

$n_e=20$. Figures 6 and 7 show the decision trees constructed in the best and worst runs for $n_e=20$ (run 3 and run 4), respectively. Comparing these two trees, we notice that two patterns that appeared at the upper levels of each tree are identical. The pattern appeared at the root node of both trees is shown in Figure 8. From the overall contingency table for this experiment shown in Table 4, it is found that, since the overall error rate in experiment 2 was larger than that of experiment 1, the number of misclassified instances increased.

The average prediction error rate in experiment 1 is better than that in experiment 2, as the difference in characteristics between data in F4 stage and data in {F0+F1} stages is intuitively larger than that between data in F4 stage and data in {F3+F2}. The averaged error rate of 12.50% in experiment 1 is fairly comparable to that of 11.8% obtained by the decision tree reported in [9]. By regarding LC (F4) as positive and non-LC ({F3+F2} in experiment 1 and {F3+F2} in experiment 2) as negative, the overall contingency table shows decision trees constructed by DT-GBI tended to have more false negative than false positive for predicting the stage of fibrosis. This tendency was more prominent in experiment 2 compared with experiment 1.

Patterns shown in Figures 5 and 8 are sufficiently discriminative since all of them are used at the nodes in the upper part of all decision trees. The credibility of these patterns is ensured as, for almost all patients, they appear after the

Table 5. Size of graphs (hepatitis type)

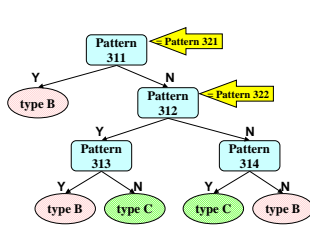
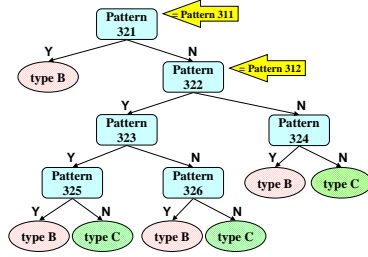
hepatitis type	Type B	Type C	Total
No. of graphs	77	185	262
Avg. No. of nodes	238	286	272
Max. No. of nodes	375	377	377
Min. No. of nodes	150	167	150

Table 7. Contingency table with the number of instances (hepatitis type)

Actual Class	Predicted Class	
	Type B	Type C
Type B	559	211
Type C	181	979

Table 6. Average error rates (%) (hepatitis type)

run of 10 CV	Type B vs. Type C	
	$n_r=20$	$n_e=20$
1	21.76	18.65
2	21.24	19.69
3	21.24	19.17
4	23.32	20.73
5	25.39	22.80
6	25.39	23.32
7	22.28	18.65
8	24.87	19.17
9	22.80	19.69
10	23.83	21.24
Average	23.21	20.31
Standard Deviation	1.53	1.57

**Fig. 9.** One of the ten trees from the best run in exp.3 ($n_e=20$)**Fig. 10.** One of the ten trees from the worst run in exp.3 ($n_e=20$)

biopsy. These patterns include inspection items and their values that are typical of cirrhosis.

4.3 Experiment 3: Classifying Patients with Types (B or C)

There are two types of hepatitis recorded in the dataset; B and C. We constructed decision trees which distinguish between patients of type B and type C with the same duration and interval as in subsection 4.2. On the other hand, we used only the 25 attributes other than those of antigen and antibody (HBC-AB, HBE-AB, HBE-AG, HBS-AB, HBS-AG, HCV-AB, HCV-RNA) because they obviously indicate the type of hepatitis. Table 5 shows the size of graphs after the data conversion. To keep the number of instances at 2:3 ratio [9], we used all of 77 instances in type B as “Type B” class and 116 instances in type C as “Type C” class. Hence, there are 193 instances in all. b was set to 5 in this experiment.

The overall result is summarized in Table 6. The average error rate was 23.21% for $n_r=20$ and 20.31% for $n_e=20$. Figures 9 and 10 show the decision trees

Table 8. Class label for interferon therapy

label	
R	virus disappeared (Response)
N	virus existed (Non-response)
?	no clue for virus activity
R?	R (not fully confirmed)
N?	N (not fully confirmed)
??	missing

Table 9. Size of graphs (interferon therapy)

effectiveness of interferon therapy	R	N	Total
No. of graphs	38	56	94
Avg. No. of nodes	77	74	75
Max. No. of nodes	123	121	123
Min. No. of nodes	41	33	33

Table 10. Average error rate (%) (interferon therapy)

run of 10 CV	$n_e=20$
1	18.65
2	19.69
3	19.17
4	20.73
5	22.80
6	23.32
7	18.65
8	19.17
9	19.69
10	21.24
Average	22.60
Standard Deviation	1.57

constructed in the best and worst runs for $n_e = 20$ (run 1 and run 6), respectively. Comparing these trees, the patterns 311 and 312 were identical to the patterns 321 and 322, respectively and appeared at the upper level nodes in almost all the decision trees. Thus they are considered sufficiently discriminative. The overall contingency table for this experiment is shown in Table 7. By regarding type B as positive and type C as negative, it shows decision trees constructed by DT-GBI tended to have more false negative than false positive, similarly to the results in experiment 1 and 2.

4.4 Experiment 4: Classifying Interferon Therapy

An interferon is a medicine to get rid of hepatitis virus and it is said that the smaller the amount of virus is, the more effective interferon therapy is. Since, unfortunately, the hepatitis dataset does not contain the examination record for the amount of virus, response to interferon therapy was judged by a medical doctor for each patient, which was used as the class label for interferon therapy. The resulting class labels are summarized in Table 8. Note that this experiment was conducted for the patients with label R (38 patients) and N (56 patients). Medical records for other patients were not used.

To analyze the effectiveness of interferon therapy, we hypothesized that the amount of virus in a patient was almost stable for a certain duration just before the interferon injection in the dataset. Data in the range of 90 days to 1 day before the administration of interferon were extracted for each patient and average was taken for two-week interval. Furthermore, we hypothesized that each pathological condition in the extracted data could directly affect the pathological condition just before the administration. To represent this dependency, each subgraph was directly linked to the last subgraph in each patient. Since the objective of this analysis is to predict the effectiveness of interferon therapy without referring to the amount of virus, we used the same 25 attributes other than those

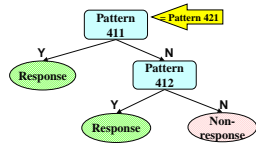


Fig. 11. One of the ten trees from the best run in exp.4

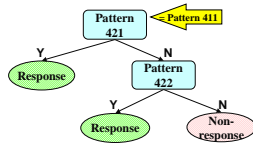


Fig. 12. One of the ten trees from the worst run in exp.4

Table 11. Contingency table with the number of instances (interferon therapy)

Actual Class	Predicted Class	
	R	N
R	250	130
N	83	477

of antigen and antibody as in subsection 4.3. Table 9 shows the size of graphs after the data conversion. b was set to 3 in experiment 4.

The results are summarized in Table 10 and the overall average error rate was 22.60% (in this experiment we did not run the cases for $n_r=20$). Figures 11 and 12 show the decision trees constructed in the best and worst runs (run 1 and run 6), respectively. Although the structure of decision tree in Figure 11 is simple, its prediction accuracy was actually good (error rate=10%). Furthermore, since the pattern 411 (identical to the pattern 421) was used at the root node of many decision trees, it is considered as sufficiently discriminative for classifying patients for whom interferon therapy was effective (with class label R).

The overall contingency table for this experiment is shown in Table 11. By regarding R (Response) as positive and N (Non-response) as negative, it shows that the resulting decision trees tended to have more false negative for predicting the effectiveness of interferon therapy.

5 Conclusion

We analyzed the hepatitis dataset by DT-GBI. Four experiments were conducted on the dataset and both constructed decision trees and their predictive accuracies were reported. DT-GBI was successful in extracting discriminative patterns, and the obtained prediction error rate results are thought satisfactory considering the various kinds of noises embedded in the data. We thus believe that DT-GBI is a useful tool for practicing evidence-based medicine. General finding is that the decision trees constructed by DT-GBI tended to misclassify more instances with minority class than with the majority class. Incorporating cost-sensitive learning might be effective to take into account the unbalance in class distribution and different penalties for misclassification error.

As future work, effectiveness of DT-GBI against the hepatitis data set with another way of preparing data should be examined, *e.g.*, estimating missing values, randomly selecting instances from non-cirrhosis class both for training and testing, etc. The validity of extracted patterns is now being evaluated and discussed by the domain experts (medical doctors).

Acknowledgment

This work was partially supported by the grant-in-aid for scientific research 1) on priority area “Realization of Active Mining in the Era of Information Flood” (No. 13131101, No. 13131206) and 2) No. 14780280 funded by the Japanese Ministry of Education, Culture, Sport, Science and Technology.

References

1. T. B. Ho, T. D. Nguyen, S. Kawasaki, S.Q. Le, D. D. Nguyen, H. Yokoi, and K. Takabayashi. Mining hepatitis data with temporal abstraction. In *Proc. of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 369–377, August 2003.
2. T. Matsuda, T. Yoshida, H. Motoda, and T. Washio. Knowledge discovery from structured data by beam-wise graph-based induction. In *Proc. of the 7th Pacific Rim International Conference on Artificial Intelligence (Springer Verlag LNAI2417)*, pages 255–264, 2002.
3. M. Ohsaki, Y. Sato, H. Yokoi, and T. Yamaguchi. A rule discovery support system for sequential medical data - in the case study of a chronic hepatitis dataset -. In *Working note of ECML/PKDD-2003 Discovery Challenge*, pages 154–165, 2003.
4. S. Tsumoto, K. Takabayashi, M. Nagira, and S. Hirano. Trend-evaluation multi-scale analysis of the hepatitis dataset. In *Project “Realization of Active Mining in the Era of Information Flood” Report*, pages 191–197, March 2003.
5. G. Warodom, T. Matsuda, T. Yoshida, H. Motoda, and T. Washio. Classifier construction by graph-based induction for graph-structured data. In *Proc. of the 7th Pacific-Asia Conference on Knowledge Discovery and Data Mining (Springer Verlag LNAI2637)*, pages 52–62, 2003.
6. G. Warodom, T. Matsuda, T. Yoshida, H. Motoda, and T. Washio. Performance evaluation of decision tree graph-based induction. In *Proc. of the 6th Pacific-Asia Conference on Discovery Science (Springer Verlag LNAI2843)*, pages 128–140, 2003.
7. G. Warodom, T. Yoshida, K. Ohara, H. Motoda, and T. Washio. Extracting diagnostic knowledge from hepatitis dataset by decision tree graph-based induction. In *Working note of International Workshop on Active Mining (AM2003)*, pages 106–117, 2003.
8. T. Watanabe, E. Suzuki, H. Yokoi, and K. Takabayashi. Application of prototype-line to chronic hepatitis data. In *Working note of ECML/PKDD-2003 Discovery Challenge*, pages 166–177, 2003.
9. Y. Yamada, E. Suzuki, H. Yokoi, and K. Takabayashi. Decision-tree induction from time-series data based on a standard-example split test. In *Proc. of the 12th International Conference on Machine Learning*, pages 840–847, August 2003.
10. K. Yoshida and H. Motoda. Clip : Concept learning from inference pattern. *Journal of Artificial Intelligence*, 75(1):63–92, 1995.
11. T. Yoshida, G. Warodom, A. Mogi, K. Ohara, H. Motoda, T. Washio, H. Yokoi, and K. Takabayashi. Preliminary analysis of interferon therapy by graph-based induction. In *Working note of International Workshop on Active Mining (AM2004)*, pages 31–40, 2004.