# Multi-Aspect Hepatitis Data Analysis

Yasuo Fujita†, Muneaki Ohshima†, Ning Zhong†, Hideto Yokoi‡
†Department of Information Engineering, Maebashi Institute of Technology,
460-1 Kamisadori-cho, Maebashi 371-0816, Japan
fujita@kis-lab.com
‡School of Medicine, Chiba University,
1-8-1 Inohana, Chuo-ku, Chiba 260-8677, Japan

***Abstract*— When therapy using IFN (interferon) medication for chronic hepatitis patients, various conceptual knowledge/rules will benefit for giving a treatment. In this paper, we describe an ongoing work on using various data mining agents including the GDT-RS inductive learning system for discovering classification rules, the LOI (learning with ordered information) for discovering important features, as well as the POM (peculiarity oriented mining) for finding peculiarity data/rules, in a multi-phase process for multi-aspect analysis of the hepatitis data. Our methodology and experimental results show that the perspective of medical doctors will be changed from a single type of experimental data analysis towards a holistic view, by using our *multi-aspect mining* approach.**

## I. Introduction

*Multi-aspect mining* in a multi-phase KDD (Knowledge Discovery and Data Mining) process is an important methodology for knowledge discovery from real-life data [2], [7], [11], [12]. There are two main reasons why a multi-aspect mining approach needs to be used for the hepatitis data analysis.

The first reason is that we cannot expect to develop a single data mining algorithm for analyzing all main aspects of the hepatitis data towards a holistic view since complexity of the real-world applications. Hence, various data mining agents need to be cooperatively used in the multi-phase data mining process for performing multi-aspect analysis as well as multi-level conceptual abstraction and learning.

The other reason is that when performing multi-aspect analysis for complex problems such as the hepatitis data mining, a data mining task needs to be decomposed into sub-tasks. Thus these sub-tasks can be solved by using one or more data mining agents that are distributed over different computers. Thus the decomposition problem leads us to the problem of distributed cooperative system design.

More specifically, when therapy using IFN (interferon) medication for chronic hepatitis patients, various conceptual knowledge/rules will benefit for giving a treatment. The knowledge/rules, for instance, include (1) when the IFN should be used for a patient so that he/she will be able to be cured, (2) what kinds of inspections are important for a diagnosis, and (3) whether some peculiar data/patterns exist or not.

In this paper, we describe an ongoing work on using various data mining agents including the GDT-RS inductive learning system for discovering classification rules [8], [13], the LOI (learning with ordered information) for discovering important features [4], [14], as well as the POM (peculiarity oriented mining) for finding peculiarity data/rules [15], for multi-aspect analysis of the hepatitis data so that such rules mentioned above can be discovered automatically.

We emphasize that both pre-processing/post-processing steps are important before/after using data mining agents. In particular, informed knowledge discovery, in general, uses background knowledge obtained from experts (e.g. medical doctors) about a domain (e.g. chronic hepatitis) to guide a discovery process with multi-phase such as pre-processing, rule mining, and post-processing, towards finding interesting and novel rules/features hidden in data. Background knowledge may be of several forms including rules already found, taxonomic relationships, causal preconditions, ordered information, and semantic categories.

In our experiments, the result of the blood test of the patients, who performed INF before starting medication, is first pre-treated. After that, the pre-processed data are used for each data mining agent, respectively. By using the GDT-RS, the rules with respect to know whether a medical treatment is effective or not, can be found. And, by using the LOI, what attributes affect the medical treatment of hepatitis $C$ greatly can be investigated. Our methodology and experimental results show that the perspective of medical doctors will be changed from a single type of experimental data analysis towards a holistic view, by using our multi-aspect mining approach.

The rest of the paper is organized as follows. Section 2 describes how to pre-process the hepatitis data and decide the threshold values for condition attributes according to the background knowledge obtained from medical doctors. Section 3 gives the main results mined by using the GDT-RS and the post-processing. Section 4 discusses the analysis and evaluation of the results given in Section 3, based on a medical doctor's advice and comments. Then in Section 5, we extend our system by adding the LOI (learning with ordered information) and POM (peculiarity oriented mining) data mining agents for multi-aspect mining and analysis. Finally, Section 6 gives conclusions.

## II. Mining by GDT-RS

### A. Pre-processing

*1) Selection of Inspection Data and Class Determination:* We use the following conditions to extract inspection data.

- Patients of chronic hepatitis type $C$ who may be medicated with IFN.

- Patients with the data of judging the IFN effect by using whether a hepatitis virus exists or not.
- Patients with inspection data collected in one year before IFN is used.

Thus, 197 patients with 11 condition attributes as shown in Table I are selected and will be used in our data mining agents.

TABLE I

CONDITION ATTRIBUTES

| | | | |
|---|---|---|---|
| T-CHO | CHE | ALB | TP |
| T-BIL | D-BIL | I-BIL | PLT |
| WBC | HGB | GPT | |

Furthermore, the decision attribute (i.e. classes) is selected according to a result of judging the IFN effect by using whether a hepatitis virus exists or not. Hence, the 197 extracted patients can be classified into 3 classes as shown in Table II.

TABLE II

THE DECISION ATTRIBUTE (CLASSES)

| class | The condition of the patient after IFN | # of patients |
|---|---|---|
| R | Disappearance of the virus | 58 |
| N | Existence of virus | 86 |
| ? | Reliability lack of data | 53 |

*2) Evaluation of Condition Attributes:* As shown in Fig. 1, the condition attributes are evaluated as follows.

1) All the inspection values in one year before IFN is used for each patient are divided into two groups, the first half and the second half of the inspection values.
2) When the absolute value of the difference between average values of the first half and the second half of the inspection values exceeds the threshold, it is estimated as up or down. Otherwise, it is estimated as "–" (i.e. no change). Moreover, it is estimated as "?" in the case where not inspection data or only once (i.e. a patient is examined only once).

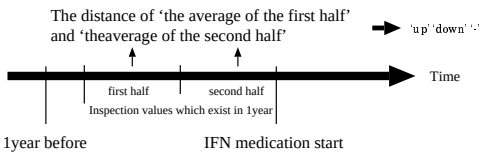


Fig. 1. The evaluation method of condition attributes

Furthermore, the threshold values can be decided as follows.

- **The threshold values for each attribute except GPT** are set up to 10% of the normal range of each inspection data. As the change of a hepatitis patient's GPT value will exceed the normal range greatly, the threshold value for the GPT needs to be calculated in a more complex method to be described below. The threshold values used for evaluating each condition attribute is shown in Table III.
- **The threshold value for GPT** is calculated as follows. As shown in Fig. 2, the standard deviation of the difference of the adjacent test values of each hepatitis patient's GPT is first calculated, respectively; And then the standard deviation of such standard deviation is used as a threshold value. In the GPT test, let $m$ be the number of patients, $t_m(1 \leq m \leq M)$ the time of test (that may be different for each patient), $d_{mi}(1 \leq i \leq t_m - 1)$ the difference of adjacent test values. Thus, the threshold value of GPT can be calculated in the following equation.

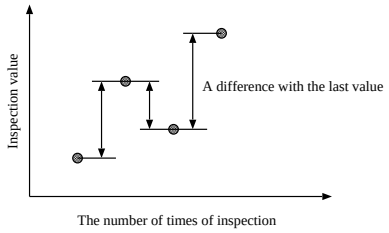


Fig. 2. Standard deviation of the difference of adjacent values

$$\text{threshold of GPT} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(s_m - \overline{s})^2} \qquad (1)$$

where $s_m(1 \leq m \leq M)$ is the standard deviation of the difference $d_{mi}$ of the test value that is calculated for each patient, respectively, and $\overline{s}$ is the average value of $s_m$.

Finally, the threshold values are set up as shown in Table III.

TABLE III

THE THRESHOLD VALUES FOR EVALUATING CONDITION ATTRIBUTES

| | | | |
|---|---|---|---|
| T-CHO $> 9.5$ | CHE $> 25$ | ALB $> 0.12$ | TP $> 0.17$ |
| T-BIL $> 0.1$ | D-BIL $> 0.03$ | I-BIL $> 0.07$ | PLT $> 20$ |
| WBC $> 0.5$ | HGB $> 0.6$ | GPT $> 54.56$ | |

### B. Main Results

*1) Rule Discovery by GDT-RS:* GDT-RS is a soft hybrid induction system for discovering classification rules from databases with uncertain and incomplete data [8], [13]. The system is based on a hybridization of the *Generalization Distribution Table (GDT)* and the *Rough Set* methodology. The main features of GDT-RS are the following:

- Biases for search control can be selected in a flexible way. Background knowledge can be used as a bias to control the initiation of GDT and in the rule discovery process.
- The rule discovery process is oriented toward inducing rules with high quality of classification of unseen instances. The rule uncertainty, including the ability to predict unseen instances, can be explicitly represented by the rule strength.
- A minimal set of rules with the minimal (semi-minimal) description length, having large strength, and covering of all instances can be generated.
- Interesting rules can be induced by selecting a discovery target and class transformation.

In the experimental results at the accuracy 60%, only the rules with which the number of condition attributes is less than

TABLE IV

RULES WITH RESPECT TO CLASS $R$

| rule-ID | rule & accuracy |
|---|---|
| 001 | GPT(up) & (10/16)=62% |
| 002 | T-CHO(down) ∧ PLT(down) & (6/9)=66% |
| 003 | T-BIL(up) ∧ GPT(down) & (3/4)=75% |
| 004 | TP(down) ∧ GPT(down) & (3/4)=75% |

TABLE V

RULES WITH RESPECT TO CLASS $N$

| rule-ID | rule & accuracy |
|---|---|
| 101 | D-BIL(down) & (26/43)=60% |
| 102 | T-CHO(down) ∧ I-BIL(down) & (7/11)=63% |
| 103 | I-BIL(down) ∧ WBC(down) & (7/8)=87% |
| 104 | D-BIL(up) ∧ PLT(down) & (4/6)=66% |
| 105 | TP(up) ∧ I-BIL(down) & (5/6)=83% |
| 106 | TP(up) ∧ T-BIL(down) & (4/6)=66% |
| 107 | TP(up) ∧ PLT(down) & (4/5)=80% |
| 108 | CHE(up) ∧ T-BIL(down) & (2/4)=50% |

or equal to three are extracted. This is because it will become unclear if the number of condition attributes increases. Tables IV and V show such rules that are divided into classes $R$ and $N$, respectively.

*2) Results of Post-processing:* As a post-processing, we checked each discovered rule covers what patient(s) related data. Table VI shows the results, where the $Positive$ (or $Negative$) $ID$ means that the patient is covered by a rule as a $positive$ (or $negative$) instance. From this table, we can see it becomes clear that what patient group is covered by what rule. Hence it is useful for finding the main features of a patient group.

TABLE VI

PATIENTS COVERED BY RULES WITH RESPECT TO CLASS $R$

| rule-ID | Positive patient ID | | | | | Negative patient ID | | | |
|---|---|---|---|---|---|---|---|---|---|
| 001 | 158 | 351 | 534 | 547 | 778 | 35 | 188 | 273 | 452 |
| | 801 | 909 | 923 | 940 | 942 | 623 | 712 | | |
| 002 | 91 | 351 | 650 | 703 | 732 | 169 | 712 | 952 | |
| | 913 | | | | | | | | |
| 003 | 431 | 592 | 700 | | | 122 | | | |
| 004 | 37 | 71 | 730 | | | 122 | | | |

As an example of post-processing, Table VIII shows a part of result of the post-processing about class $R$. Here "+" and "–" denote the patient covered by a rule as a positive or negative instance, respectively. For example, *rule 001* is covered by the patient IDs: {158, 778, 801, 909, 923, 940, 942}.

## C. Analyses and Evaluations

The results derived by the GDT-RS and post-processing have been evaluated by a medical doctor based on acceptability and novelty of each rule. The evaluations of the rules are divided into five stages: 1 is the lowest and 5 is the highest evaluation for acceptability and novelty of each rule.

TABLE VII

PATIENTS COVERED BY RULES WITH RESPECT TO CLASS $N$

| rule-ID | Positive patient ID | | | | | Negative patient ID | | | |
|---|---|---|---|---|---|---|---|---|---|
| 101 | 2 | 104 | 125 | 182 | 184 | 37 | 71 | 133 | 169 |
| | 191 | 203 | 208 | 239 | 290 | 180 | 206 | 248 | 276 |
| | 546 | 439 | 493 | 498 | 529 | 413 | 593 | 610 | 683 |
| | 578 | 585 | 634 | 652 | 653 | 702 | 713 | 732 | 771 |
| | 669 | 715 | 719 | 743 | 750 | 948 | | | |
| | 756 | | | | | | | | |
| 102 | 2 | 239 | 563 | 634 | 652 | 169 | 413 | 650 | 732 |
| | 653 | 952 | | | | | | | |
| 103 | 2 | 138 | 208 | 432 | 578 | 413 | | | |
| | 653 | 736 | | | | | | | |
| 104 | 187 | 260 | 289 | 712 | | 703 | 778 | | |
| 105 | 72 | 182 | 219 | 546 | | 35 | | | |
| | 920 | | | | | | | | |
| 106 | 72 | 182 | 219 | 546 | | 180 | 610 | | |
| 107 | 104 | 182 | 260 | 535 | | 180 | | | |
| 108 | 210 | 634 | | | | 180 | 683 | | |

TABLE VIII

POST-PROCESSING ABOUT CLASS $R$

| Patient ID | rule 001 | rule 002 | rule 003 | rule 004 |
|---|---|---|---|---|
| 35 | – | | | |
| 37 | | | | + |
| 71 | | | | + |
| 78 | | | | |
| 91 | | + | | |
| 122 | | | – | – |
| 158 | + | | | |
| 169 | | – | | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 778 | + | | | |
| 801 | + | | | |
| 909 | + | | | |
| 913 | | + | | |
| 923 | + | | | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 940 | + | | | |
| 942 | + | | | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

*1) Evaluation of Rules:* From the viewpoint of the rules with a higher support (e.g. *rule-001* and *rule-101*), we observed that

- It will heal up in many cases if a patient is medicated with IFN at the time when GPT is going up (hepatitis is getting worse);
- It does not heal up in many cases even if a patient is medicated with IFN at the time when D-BIL is descending.

Furthermore, the following two points on the effect of IFN is understood clearly.

- It is relevant to different types of hepatitis viruses;
- It is hard to be effective when there are large amounts of hepatitis virus.

Hence, we can see that *rule-001* and *rule-101* do not conflict with the existing medicine knowledge.

From the two rules, the hypothesis: "IFN is more effective when the inflammation of hepatitis is stronger" can be formed. Based on this hypothesis, we can evaluate the rules discovered as follows.

- In class *R*, the acceptability of the rules with respect to aggravation of liver function will be good.

TABLE IX

EVALUATION OF RULES WITH RESPECT TO CLASS $R$

| rule-ID | acceptability | novelty |
|---|---|---|
| 001 | 4 | 5 |
| 002 | 3 | 5 |
| 003 | 4 | 5 |
| 004 | 4 | 5 |

TABLE X

EVALUATION OF RULES WITH RESPECT TO CLASS $N$

| rule-ID | acceptability | novelty |
|---|---|---|
| 101 | 4 | 5 |
| 102 | 2 | 3 |
| 103 | 2 | 3 |
| 104 | 1 | 1 |
| 105 | 3 | 4 |
| 106 | 3 | 4 |
| 107 | 2 | 3 |
| 108 | 3 | 4 |

- In class *N*, the acceptability of the rules with respect to recovery of liver function will be good.

Hence, the evaluations as shown in Tables IX and X can be obtained. In class $N$, we can see that the acceptability of some rules is 2. This is because both the recovery and aggravation of liver function are included in the premise of the rules.

*2) Evaluation of Post-processing:* In the discovered rules, we can see there is some relevance among the patients supported by bilirubin (T-BIL, D-BIL, I-BIL) in class $N$. From the relation denoted in T-BIL = D-BIL + I-BIL, it is clear that the rules with respect to bilirubin are relevant. Hence the rules are supporting the same patients group.

Moreover, in order to examine the hypothesis, "the medical background which a rule shows is not contradictory to a patient's condition", the discovered rules are categorized, based on liver function, into three categories: recovery, aggravation, or mixed recovery and aggravation, as shown in Table XI.

From Table XI, we observed that there are many rules with the same conditions in the rule group supported by a patients group, and it may conflict with unknown medical background that is not represented in the conditions of the rules. However, it does not mean that the rules are incorrect. The reason may be that the rules cannot be simply categorized by recovery and aggravation.

For example, although it can show liver function aggravation, the lower values of WBC and ALB may not be the real reason

TABLE XI

CATEGORY OF DISCOVERED RULES

| | Recovery | Aggravation | Recovery & Agg. |
|---|---|---|---|
| class R | rule 007<br>rule 008<br>rule 009<br>rule 011 | rule 001<br>rule 002 | rule 003<br>rule 004<br>rule 005<br>rule 006<br>rule 010 |
| class N | rule 101<br>rule 105<br>rule 106<br>rule 108<br>rule 110 | rule 104<br>rule 109 | rule 102<br>rule 103<br>rule 107<br>rule 111 |

of liver function aggravation. On other hand, since WBC and PLT are the same blood cell ingredient, and T-CHO and ALB are relevant to protein that makes liver, they may be relevant from this point of view. However, T-CHO and ALB do not only provide for liver, but also, for example, T-CHO is related to eating, and ALB is related to the kidney, respectively. Hence it cannot declare there is such correlation.

In summary, there is correlation if we are mentioning about mathematical relevance like BIL. However, it is difficult to find out correlation for others. We need the following methods to solve the issue.

- Finding out what rules are significant from the statistical point of view, based on rough categorizing such as recovery and aggravation.
- Showing whether such rough categorizing is sufficient or not.

## III. MULTI-ASPECT ANALYSIS BY LOI AND POM

Based on the results stated above, we have been extending our system by adding the LOI (learning with ordered information) and POM (peculiarity oriented mining) data mining agents for multi-aspect mining and analysis.

### A. Mining by LOI

The LOI uses background knowledge called *ordered relation* for discovering *ordering rules* and important attributes for an ordered decision class [4], [14]. For example, since the values for T-CHO are the larger and the better, the ordered relation can be set to (VH $\succ$ H $\succ$ N $\succ$ L $\succ$ VL), where "$\succ$" denotes a weak order. Furthermore, if a decision attribute has two classes: $R$ (response) and $N$ (no response), the ordered relation can be set to $R \succ N$.

An ordered information table may be viewed as information tables with added semantics (background knowledge). The following figure shows an example of creating such a table in which background knowledge is included.

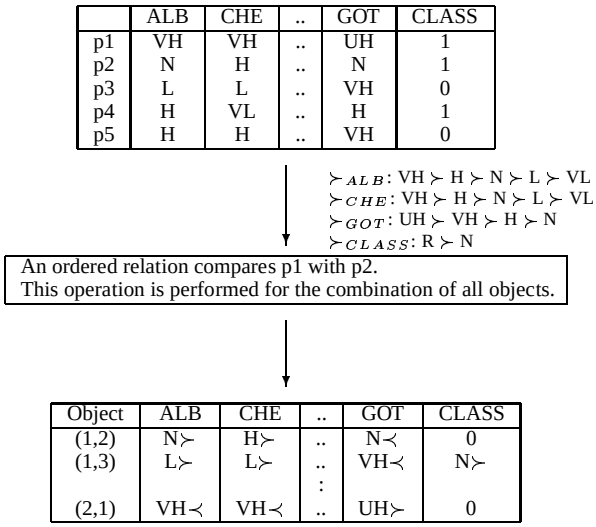


Fig. 3. Creating an ordered information table

After this transformation, ordering rules can be discoverd from the ordered information table by our GDT-RS rule mining

TABLE XII

DISCOVERED RULES BY LOI

| Rules(SupportNUM 10, Support 70%) | Support |
|---|---|
| PLT(N$\prec$) $\wedge$ DBIL(H$\succ$) $\wedge$ GPT(VH$\succ$) | 27/30=90% |
| WBC(L$\succ$) $\wedge$ GOT(VH$\succ$) $\wedge$ GPT(VH$\succ$) | 20/22=90% |
| PLT(VL$\succ$) $\wedge$ GOT(VH$\succ$) $\wedge$ GPT(VH$\succ$) | 16/18=88% |
| PLT(L$\succ$) $\wedge$ TP(H$\prec$) $\wedge$ GOT(VH$\prec$) | 15/16=93% |
| DBIL(H$\succ$) $\wedge$ GPT(VH$\succ$) | 14/17=82% |
| HGB(N$\prec$) $\wedge$ WBC(L$\succ$) $\wedge$ GOT(VH$\succ$) | 16/16=100% |
| DBIL(H$\succ$) $\wedge$ GOT(H$\succ$) | 14/16=87% |
| DBIL(H$\succ$) $\wedge$ GOT(H$\succ$) $\wedge$ GPT(VH$\succ$) | 12/15=80% |
| ALB(L$\prec$) $\wedge$ WBC(L$\succ$) $\wedge$ GPT(VH$\succ$) | 10/12=83% |

TABLE XIII

THE CORRECTION RATE OF THE BACKGROUND KNOWLEDGE

| Attribute | $att_{\rm pos}$ | $att_{\rm neg}$ |
|---|---|---|
| ALB | 58.9% | 67.2% |
| CHE | 54.7% | 70.6% |
| D-BIL | 26.8% | 35.3% |
| GOT | 34.2% | 47.3% |
| GPT | 36.7% | 47.8% |
| HGB | 78.0% | 81.4% |
| I-BIL | 32.5% | 53.1% |
| PLT | 36.6% | 49.5% |
| T-BIL | 44.0% | 47.1% |
| T-CHO | 30.6% | 48.3% |
| TP | 62.7% | 78.3% |
| WBC | 9.5% | 30.5% |

system. The rules discovered by LOI are shown in Table XII. From this table, we can see that condition attributes included in the rules can be divided into two types: "go better" and "go worse" corresponding to the class vaules: chronic hepatitis patients to be cured or not.

The evaluation of acceptability and novelty for the rules depends to a great extent on the background knowledge with respect to the ordered information is correct or not. By investigating the values of each attribute in the ordered information table by using Eqs. (2) and (3), the correction rate of the background knowledge with respect to "go better" (or "go worse") can be obtained as shown in Table XIII.

$$att_{\rm pos} = \frac{\#ATT_{\succ,\succ}}{\#ATT_{\succ,\succ} + \#ATT_{\succ,\prec}} \quad (2)$$

$$att_{\rm neg} = \frac{\#ATT_{\prec,\prec}}{\#ATT_{\prec,\succ} + \#ATT_{\prec,\prec}} \quad (3)$$

where $\#ATT$ is the number of different values of each attribute in the ordered information table, $x = CLASS(i,j)$, $y = ATT(i,j)$, and $(i,j) \in Object$. As shown in Table XIII, the background knowledge with a high correction rate (e.g. 'TP = 78.3%'and 'HGB = 81.4%') can be explained that the background knowledge is consistent with the specific characteristics of the real collected data. On the contrary, the background knowledge with a very low correction rate (e.g. 'WBC = 9.5%') may mean that the ordered information given by an expert may not suitable for the specific data analysis.

One explanation for this situation is that the ordered information given by an expert may be too general and dose not meet specific characteristics of the real collected data, although the background knowledge is correct in general. In this case, the ordered information as common background knowledge needs to be adjusted according to specific characteristics of the real data such as the distribution and clusters of the real data. In other words, different background knowledge needs to be used for different special situations. How to adjust the ordered information is an important ongoing work in this research direction.

### *B. Mining by POM (Peculiarity Oriented Mining)*

Peculiarity represents a new interpretation of interestingness, an important notion long identified in data mining [3], [10], [15]. Peculiarity, unexpected relationships/rules may be hidden in a relatively small number of data. *Peculiarity rules* are a typical regularity hidden in many scientific, statistical, medical, and transaction databases. They may be difficult to find by applying the standard association rule mining method [1], due to the requirement of large support. In contrast, the POM (peculiarity oriented mining) agent focuses on some interesting data (peculiar data) in order to find novel and interesting rules (peculiarity rules).

We have been applying the POM for hepatitis data analysis and had some preliminary result [6]. Currently, we are also working with Suzuki's group to integrate the Peculiarity Oriented Mining approach with the Exception Rules/Data Mining approach for discovering more refined LC (Liver Cirrhosis) and non-LC classification models. The ongoing work will be also reported in detail in our next papers.

## IV. CONCLUSIONS

We presented a multi-aspect mining approach in a multi-phase, multi-aspect hepatitis data analysis process. Both pre-processing and post-processing steps are important before/after using data mining agents. Informed knowledge discovery in real-life hepatitis data needs to use background knowledge obtained from medical docotors to guide the discovery process with multi-phase such as pre-processing, rule mining, and post-processing, towards finding interesting and novel rules/features hidden in data.

Our methodology and experimental results show that the perspective of medical doctors will be changed from a single type of experimental data analysis towards a holistic view, by using our multi-aspect mining approach in which various data mining agents are used in a distributed cooperative mode in the spiral discovery process.

## REFERENCES


[1] Agrawal R. et al. "Fast Discovery of Association Rules", *Advances in Knowledge Discovery and Data Mining*, AAAI Press (1996) 307-328.

[2] Fayyad, U.M., Piatetsky-Shapiro, G, and Smyth, P. "From Data Mining to Knowledge Discovery: an Overview", *Advances in Knowledge Discovery and Data Mining*, MIT Press (1996) 1-36.

[3] Liu, B., Hsu W., Chen, S., and Ma, Y. "Analyzing the Subjective Interestingness of Association Rules", *IEEE Intelligent Systems*, Vol.15, No.5 (2000) 47-55.

[4] Sai, Y., Yao, Y.Y., and Zhong, N. "Data Analysis and Mining in Ordered Information Tables", *Proc. 2001 IEEE International Conference on Data Mining (ICDM'01)*, IEEE Computer Society Press (2001) 497-504.

[5] Suzuki, E. Undirected Discovery of Interesting Exception Rules, *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific, Vol.16, No.8 (2002) 1065-1086.

[6] Yokoi, H., Hirano, S., Takabayashi, K., Tsumoto, S., and Satomura, Y. "Active Mining in Medicine: A Chronic Hepatitis Case – Towards Knowledge Discovery in Hospital Information Systems", *Journal of Japanese Society for Artificial Intelligence*, Vol.17, No.5 (2002) 622-628 (in Japanese).

[7] Zhong, N. and Ohsuga, S. "Toward A Multi-Strategy and Cooperative Discovery System", *Proc. First Int. Conf. on Knowledge Discovery and Data Mining (KDD-95)*, AAAI Press (1995) 337-342.

[8] Zhong, N., Dong, J.Z., and Ohsuga, S., "Rule Discovery in Medical Data by GDT-RS", Special Issue on Comparison and Evaluation of KDD Methods with Common Medical Datasets, *Journal of Japanese Society for Artificial Intelligence*, Vol.15, No.5 (2000) 774-781 (in Japanese).

[9] Zhong, N. "Knowledge Discovery and Data Mining", *the Encyclopedia of Microcomputers*, Volume 27 (Supplement 6) Marcel Dekker (2001) 93-122.

[10] Zhong, N., Yao, Y.Y., Ohshima, M., and Ohsuga, S. "Interestingness, Peculiarity, and Multi-Database Mining", *Proc. 2001 IEEE International Conference on Data Mining (ICDM'01)*, IEEE Computer Society Press (2001) 566-573.

[11] Zhong, N. and Ohsuga, S. "Automatic Knowledge Discovery in Larger Scale Knowledge-Data Bases", C. Leondes (ed.) *The Handbook of Expert Systems*, Vol.4: Chapter 29, Academic Press (2001) 1015-1070.

[12] Zhong, N., Liu, C., and Ohsuga, S. "Dynamically Organizing KDD Process", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 15, No. 3, World Scientific (2001) 451-473.

[13] Zhong, N., Dong, J.Z., and Ohsuga, S. "Rule Discovery by Soft Induction Techniques", *Neurocomputing*, An International Journal, Vol. 36 (1-4) Elsevier (2001) 171-204.

[14] Zhong, N., Yao, Y.Y., Dong, J.Z., Ohsuga, S., "Gastric Cancer Data Mining with Ordered Information", J.J. Alpigini et al (eds.) *Rough Sets and Current Trends in Computing*, LNAI 2475, Springer (2002) 467-478.

[15] Zhong, N., Yao, Y.Y., Ohshima M., "Peculiarity Oriented Multidatabase Mining", *IEEE Transactions on Knowledge and Data Engineering*, Vol.15, No.4 (2003) 952-960.