Mining Positive and Negative Knowledge in Clinical Databases Based on Rough Set Model

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. One of the most important problems on rule induction methods is that extracted rules partially represent information on experts' decision processes, which makes rule interpretation by domain experts difficult. In order to solve this problem, the characteristics of medical reasoning is discussed positive and negative rules are introduced which model medical experts' rules. Then, for induction of positive and negative rules, two search algorithms are provided. The proposed rule induction method was evaluated on medical databases, the experimental results of which show that induced rules correctly represented experts' knowledge and several interesting patterns were discovered.

1 Introduction

Rule induction methods are classified into two categories, induction of deterministic rules and probabilistic ones [1,2,3,4]. On one hand, Deterministic rules are described as *if-then* rules, which can be viewed as propositions. From the set-theoretical point of view, a set of examples supporting the conditional part of a deterministic rule, denoted by C, is a subset of a set whose examples belongs to the consequence part, denoted by D. That is, the relation $C \subseteq D$ holds and deterministic rules are supported only by positive examples in a dataset. On the other hand, probabilistic rules are if-then rules with probabilistic information [4]. From the set-theoretical point of view, C is not a subset, but closely overlapped with D. That is, the relations $C \cap D \neq \phi$ and $|C \cap D|/|C| \geq \delta$ will hold in this case.[1] Thus, probabilistic rules are supported by a large number of positive examples and a small number of negative examples. The common feature of both deterministic and probabilistic rules is that they will deduce their consequence positively if an example satisfies their conditional parts. We call the reasoning by these rules *positive reasoning*.

However, medical experts do not use only positive reasoning but also *negative reasoning* for selection of candidates, which is represented as if-then rules whose consequences include negative terms. For example, when a patient who complains of headache does not have a throbbing pain, migraine should not be

---

[1] The threshold δ is the degree of the closeness of overlapping sets, which will be given by domain experts. For more information, please refer to Sect. 3.

suspected with a high probability. Thus, negative reasoning also plays an important role in cutting the search space of a differential diagnosis process [4]. Thus, medical reasoning includes both positive and negative reasoning, though conventional rule induction methods do not reflect this aspect. This is one of the reasons why medical experts have difficulties in interpreting induced rules and the interpretation of rules for a discovery procedure does not easily proceed. Therefore, negative rules should be induced from databases in order not only to induce rules reflecting experts' decision processes, but also to induce rules which will be easier for domain experts to interpret, both of which are important to enhance the discovery process done by the corporation of medical experts and computers.

In this paper, first, the characteristics of medical reasoning are focused on and two kinds of rules, positive rules and negative rules are introduced as a model of medical reasoning. Interestingly, from the set-theoretical point of view, sets of examples supporting both rules correspond to the lower and upper approximation in rough sets [2]. On the other hand, from the viewpoint of propositional logic, both positive and negative rules are defined as classical propositions, or deterministic rules with two probabilistic measures, classification accuracy and coverage. Second, two algorithms for induction of positive and negative rules are introduced, defined as search procedures by using accuracy and coverage as evaluation indices. Finally, the proposed method was evaluated on several medical databases, the experimental results of which show that induced rules correctly represented experts' knowledge and several interesting patterns were discovered.

2   Focusing Mechanism

One of the characteristics in medical reasoning is a focusing mechanism, which is used to select the final diagnosis from many candidates [4]. For example, in differential diagnosis of headache, more than 60 diseases will be checked by present history, physical examinations and laboratory examinations. In diagnostic procedures, a candidate is excluded if a symptom necessary to diagnose is not observed.

This style of reasoning consists of the following two kinds of reasoning processes: exclusive reasoning and inclusive reasoning.[2] The diagnostic procedure will proceed as follows: first, exclusive reasoning excludes a disease from candidates when a patient does not have a symptom which is necessary to diagnose that disease. Secondly, inclusive reasoning suspects a disease in the output of the exclusive process when a patient has symptoms specific to a disease. These two steps are modelled as usage of two kinds of rules, negative rules (or exclusive rules) and positive rules, the former of which corresponds to exclusive reasoning and the latter of which corresponds to inclusive reasoning. In the next two subsections, these two rules are represented as special kinds of probabilistic rules.

---

[2] Relations this diagnostic model with another diagnostic model are discussed in [5].

3   Definition of Rules

3.1   Rough Sets

In the following sections, we use the following notations introduced by Grzymala-Busse and Skowron [6], which are based on rough set theory [2]. These notations are illustrated by a small database shown in Table 1, collecting the patients who complained of headache.

Table 1. An example of database

No.	age	location	nature	prodrome	nausea	M1	class
1	50-59	occular	persistent	no	no	yes	m.c.h.
2	40-49	whole	persistent	no	no	yes	m.c.h.
3	40-49	lateral	throbbing	no	yes	no	migra
4	40-49	whole	throbbing	yes	yes	no	migra
5	40-49	whole	radiating	no	no	yes	m.c.h.
6	50-59	whole	persistent	no	yes	yes	psycho

DEFINITIONS. M1: tenderness of M1, m.c.h.: muscle contraction headache, migra: migraine, psycho: psychological pain.

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, $A = (U, A \cup \{d\})$. For example, Table 1 is an information system with $U = \{1, 2, 3, 4, 5, 6\}$ and $A = \{age, location, nature, prodrome, nausea, M1\}$ and $d = class$. For $location \in A$, $V_{location}$ is defined as $\{occular, lateral, whole\}$.

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 example, $[location = occular]$ is a descriptor of B.

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. $(f \wedge g)_A = f_A \cap g_A$; $(f \vee g)_A = f_A \vee g_A$; $(\neg f)_A = U - f_a$

For example, $f = [location = whole]$ and $f_A = \{2, 4, 5, 6\}$. As an example of a conjunctive formula, $g = [location = whole] \wedge [nausea = no]$ is a descriptor of U and f_A is equal to $g_{location, nausea} = \{2, 5\}$.

3.2 Classification Accuracy and Coverage

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

Definition 1.
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 $R \to d$ is defined as:

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

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

where $|S|$, $\alpha_R(D)$, $\kappa_R(D)$ 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.

In the above example, when R and D are set to $[nau = 1]$ and $[class = migraine]$, $\alpha_R(D) = 2/3 = 0.67$ and $\kappa_R(D) = 2/2 = 1.0$.

It is notable that $\alpha_R(D)$ measures the degree of the sufficiency of a proposition, $R \to D$, and that $\kappa_R(D)$ measures the degree of its necessity. For example, if $\alpha_R(D)$ is equal to 1.0, then $R \to D$ is true. On the other hand, if $\kappa_R(D)$ is equal to 1.0, then $D \to R$ is true. Thus, if both measures are 1.0, then $R \leftrightarrow D$.

3.3 Probabilistic Rules

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

$$R \overset{\alpha,\kappa}{\to} d \quad s.t. \ R = \wedge_j[a_j = v_k], \alpha_R(D) \ge \delta_\alpha$$
$$\text{and } \kappa_R(D) \ge \delta_\kappa,$$

This rule is a kind of probabilistic proposition with two statistical measures, which is an extension of Ziarko's variable precision model(VPRS) [9].[3]

It is also notable that both a positive rule and a negative rule are defined as special cases of this rule, as shown in the next subsections.

3.4 Positive Rules

A positive rule is defined as a rule supported by only positive examples, the classification accuracy of which is equal to 1.0. It is notable that the set supporting this rule corresponds to a subset of the lower approximation of a target concept, which is introduced in rough sets [2]. Thus, a positive rule is represented as:

$$R \to d \quad s.t. \qquad R = \wedge_j[a_j = v_k], \quad \alpha_R(D) = 1.0$$

---

[3] This probabilistic rule is also a kind of *Rough Modus Ponens* [10].

In the above example, one positive rule of "m.c.h." (muscle contraction headache) is:

$$[nausea = no] \rightarrow m.c.h. \quad \alpha = 3/3 = 1.0.$$

This positive rule is often called a deterministic rule. However, in this paper, we use a term, positive (deterministic) rules, because a deterministic rule which is supported only by negative examples, called a negative rule, is introduced as in the next subsection.

3.5   Negative Rules

Before defining a negative rule, let us first introduce an exclusive rule, the contrapositive of a negative rule [4]. An exclusive rule is defined as a rule supported by all the positive examples, the coverage of which is equal to 1.0.[4] It is notable that the set supporting an exclusive rule corresponds to the upper approximation of a target concept, which is introduced in rough sets [2]. Thus, an exclusive rule is represented as:

$$R \rightarrow d \quad s.t. \qquad R = \vee_j [a_j = v_k], \quad \kappa_R(D) = 1.0.$$

In the above example, the exclusive rule of "m.c.h." is:

$$[M1 = yes] \vee [nau = no] \rightarrow m.c.h. \quad \kappa = 1.0,$$

From the viewpoint of propositional logic, an exclusive rule should be represented as:

$$d \rightarrow \vee_j [a_j = v_k],$$

because the condition of an exclusive rule corresponds to the necessity condition of conclusion d. Thus, it is easy to see that a negative rule is defined as the contrapositive of an exclusive rule:

$$\wedge_j \neg [a_j = v_k] \rightarrow \neg d,$$

which means that if a case does not satisfy any attribute value pairs in the condition of a negative rules, then we can exclude a decision d from candidates. For example, the negative rule of m.c.h. is:

$$\neg [M1 = yes] \wedge \neg [nausea = no] \rightarrow \neg m.c.h.$$

In summary, a negative rule is defined as:

$$\wedge_j \neg [a_j = v_k] \rightarrow \neg d \quad s.t. \quad \forall [a_j = v_k] \, \kappa_{[a_j = v_k]}(D) = 1.0,$$

where D denotes a set of samples which belong to a class d.

Negative rules should be also included in a category of deterministic rules, since their coverage, a measure of negative concepts is equal to 1.0. It is also notable that the set supporting a negative rule corresponds to a subset of negative region, which is introduced in rough sets [2].

---

[4] An exclusive rule represents the necessity condition of a decision.

4 Algorithms for Rule Induction

The contrapositive of a negative rule, an exclusive rule is induced as an exclusive rule by the modification of the algorithm introduced in PRIMEROSE-REX [4], as shown in Fig. 1. This algorithm will work as follows. (1)First, it selects a descriptor $[a_i = v_j]$ from the list of attribute-value pairs, denoted by L. (2) Then, it checks whether this descriptor overlaps with a set of positive examples, denoted by D. (3) If so, this descriptor is included into a list of candidates for positive rules and the algorithm checks whether its coverage is equal to 1.0 or not. If the coverage is equal to 1.0, then this descriptor is added to $R_e r$, the formula for the conditional part of the exclusive rule of D. (4) Then, $[a_i = v_j]$ is deleted from the list L. This procedure, from (1) to (4) will continue unless L is empty. (5) Finally, when L is empty, this algorithm generates negative rules by taking the contrapositive of induced exclusive rules.

On the other hand, positive rules are induced as inclusive rules by the algorithm introduced in PRIMEROSE-REX [4], as shown in Fig. 2. For induction of positive rules, the threshold of accuracy and coverage is set to 1.0 and 0.0, respectively.

This algorithm works in the following way. (1) First, it substitutes L_1, which denotes a list of formula composed of only one descriptor, with the list L_{er} generated by the former algorithm shown in Fig. 1. (2) Then, until L_1 becomes empty, the following procedures will continue: (a) A formula $[a_i = v_j]$ is removed from L_1. (b) Then, the algorithm checks whether $\alpha_R(D)$ is larger than the threshold or not. (For induction of positive rules, this is equal to checking whether $\alpha_R(D)$ is equal to 1.0 or not.) If so, then this formula is included a list of the conditional part of positive rules. Otherwise, it will be included into M, which is used for making conjunction. (3) When L_1 is empty, the next list L_2 is generated from the list M.

5 Experimental Results

For experimental evaluation, a new system, called PRIMEROSE-REX2 (Probabilistic Rule Induction Method for Rules of Expert System ver 2.0), was developed, where the algorithms discussed in Sect. 4 were implemented.

PRIMEROSE-REX2 was applied to the following three medical domains: headache(RHINOS domain), whose training samples consist of 52119 samples, 45 classes and 147 attributes, cerebulovasular diseases(CVD), whose training samples consist of 7620 samples, 22 classes and 285 attributes, and meningitis, whose training samples consists of 1211 samples, 4 classes and 41 attributes (Table 2).

For evaluation, we used the following two types of experiments. One experiment was to evaluate the predictive accuracy by using the cross-validation method, which is often used in the machine learning literature [7]. The other experiment was to evaluate induced rules by medical experts and to check whether these rules led to a new discovery.

procedure *Exclusive and Negative Rules*;
  var
    $L : List$;
      /* A list of elementary attribute-value pairs */
  begin
    $L := P_0$;
    /* P_0: A list of elementary attribute-value pairs given in a database */
    while $(L \neq \{\})$ **do**
      begin
        Select one pair $[a_i = v_j]$ from L;
        if $([a_i = v_j]_A \cap D \neq \phi)$ **then do** /* D: positive examples of a target class d */
        begin
        $L_{ir} := L_{ir} + [a_i = v_j]$;   /* Candidates for Positive Rules */
        if $(\kappa_{[a_i=v_j]}(D) = 1.0)$
        then $R_{er} := R_{er} \wedge [a_i = v_j]$;
          /* Include $[a_i = v_j]$ into the formula of Exclusive Rule */
        end
      $L := L - [a_i = v_j]$;
      end
    Construct Negative Rules:
      Take the contrapositive of R_{er}.
  end {*Exclusive and Negative Rules*};

Fig. 1. Induction of exclusive and negative rules

Table 2. Databases

Domain	Samples	Classes	Attributes
Headache	52119	45	147
CVD	7620	22	285
Meningitis	1211	4	41

5.1 Performance of Rules Obtained

For comparison of performance, The experiments were performed by the following four procedures. First, rules were acquired manually from experts. Second, the datasets were randomly splits into new training samples and new test samples. Third, PRIMEROSE-REX2, conventional rule induction methods, AQ15 [1] and C4.5 [3] were applied to the new training samples for rule generation. Fourth, the induced rules and rules acquired from experts were tested by the new test samples. The second to fourth were repeated for 100 times and average all the classification accuracy over 100 trials. This process is a variant of repeated 2-fold cross-validation, introduced in [4].

    Experimental results(performance) are shown in Table 3. The first and second row show the results obtained by using PRIMROSE-REX2: the results in the first row were derived by using both positive and negative rules and those in the second row were derived by only positive rules. The third row shows the results

procedure *Positive Rules*;
  var
    $i : integer$; $M, L_i : List$;
  begin
    $L_1 := L_{ir}$;
    /* L_{ir}: A list of candidates generated by induction of exclusive rules */
    $i := 1$; $M := \{\}$;
    for $i := 1$ **to** n **do**
    /* n: Total number of attributes given
      in a database */
      begin
        while ($L_i \neq \{\}$) **do**
          begin
            Select one pair $R = \wedge[a_i = v_j]$ from L_i;
            $L_i := L_i - \{R\}$;
            if $(\alpha_R(D) > \delta_\alpha)$
              then **do** $S_{ir} := S_{ir} + \{R\}$;
          /* Include R in a list of the Positive Rules */
            else $M := M + \{R\}$;
          end
        $L_{i+1} := $ (A list of the whole combination of the conjunction formulae in M);
      end
  end {*Positive Rules*};

Fig. 2. Induction of positive rules

derived from medical experts. For comparison, we compare the classification accuracy of C4.5 and AQ-15, which is shown in the fourth and the fifth row. These

Table 3. Experimental results (accuracy: averaged)

Method	Headache	CVD	Meningitis
PRIMEROSE-REX2 (Positive+Negative)	91.3%	89.3%	92.5%
PRIMEROSE-REX2 (Positive)	68.3%	71.3%	74.5%
Experts	95.0%	92.9%	93.2%
C4.5	85.8%	79.7%	81.4%
AQ15	86.2%	78.9%	82.5%

results show that the combination of positive and negative rules outperforms positive rules, although it is a little worse than medical experts' rules.

5.2   What Is Discovered?

Positive Rules in Meningitis. In the domain of meningitis, the following positive rules, which medical experts do not expect, are obtained.

$$[WBC < 12000] \wedge [Sex = Female] \wedge [Age < 40]$$
$$\wedge [CSF\_CELL < 1000] \rightarrow Virus$$
$$[Age \geq 40] \wedge [WBC \geq 8000] \wedge [Sex = Male]$$
$$\wedge [CSF\_CELL \geq 1000] \rightarrow Bacteria$$

The former rule means that if WBC(White Blood Cell Count) is less than 12000, the Sex of a patient is FEMALE, the Age is less than 40 and CSF_CELL (Cell count of Cerebulospinal Fluid), then the type of meningitis is Virus. The latter one means that the Age of a patient is less than 40, WBC is larger than 8000, the Sex is Male, and CSF_CELL is larger than 1000, then the type of meningitis is Bacteria.

The most interesting points are that these rules included information about age and sex, which often seems to be unimportant attributes for differential diagnosis of meningitis. The first discovery was that women did not often suffer from bacterial infection, compared with men, since such relationships between sex and meningitis has not been discussed in medical context [11]. By the close examination of the database of meningitis, it was found that most of the above patients suffered from chronic diseases, such as DM, LC, and sinusitis, which are the risk factors of bacterial meningitis. The second discovery was that $[age < 40]$ was also an important factor not to suspect viral meningitis, which also matches the fact that most old people suffer from chronic diseases.

These results were also re-evaluated in medical practice. Recently, the above two rules were checked by additional 21 cases who suffered from meningitis (15 cases: viral and 6 cases: bacterial meningitis.) Surprisingly, the above rules misclassified only three cases (two are viral, and the other is bacterial), that is, the total accuracy was equal to $18/21 = 85.7\%$ and the accuracies for viral and bacterial meningitis were equal to $13/15 = 86.7\%$ and $5/6 = 83.3\%$. The reasons of misclassification were the following: a case of bacterial infection was a patient who had a severe immunodeficiency, although he is very young. Two cases of viral infection were patients who also suffered from herpes zoster. It is notable that even those misclassification cases could be explained from the viewpoint of the immunodeficiency: that is, it was confirmed that immunodeficiency is a key word for meningitis.

The validation of these rules is still ongoing, which will be reported in the near future.

Positive and Negative Rules in CVD. Concerning the database on CVD, several interesting rules were derived. The most interesting results were the following positive and negative rules for thalamus hemorrhage:

$$[Sex = Female] \wedge [Hemiparesis = Left]$$
$$\wedge [LOC : positive] \rightarrow Thalamus$$
$$\neg [Risk : Hypertension] \wedge \neg [Sensory = no]$$
$$\rightarrow \neg Thalamus$$

The former rule means that if the Sex of a patient is female and he/she suffered from the left hemiparesis([Hemiparesis=Left]) and loss of consciousness

([LOC:positive]), then the focus of CVD is Thalamus. The latter rule means that if he/she neither suffers from hypertension ([Risk: Hypertension]) nor suffers from sensory disturbance([Sensory=no]), then the focus of CVD is Thalamus.

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

5.3   Rule Discovery as Knowledge Acquisition

Expert System: RH. Another point of discovery of rules is automated knowledge acquisition from databases. Knowledge acquisition is referred to as a bottleneck problem in development of expert systems [12], which has not fully been solved and is expected to be solved by induction of rules from databases. However, there are few papers which discusses the evaluation of discovered rules from the viewpoint of knowledge acquisition [13].

For this purpose, we have developed an expert system, called RH(Rule-based system for Headache) by using the acquired knowledge.[5] RH consists of two parts. Firstly, RH requires inputs and applies exclusive and negative rules to select candidates (focusing mechanism). Then, RH requires additional inputs and applies positive rules for differential diagnosis between selected candidates. Finally, RH outputs diagnostic conclusions.

Evaluation of RH. RH was evaluated in clinical practice with respect to its classification accuracy by using 930 patients who came to the outpatient clinic after the development of this system. Experimental results about classification accuracy are shown in Table 4. The first and second row show the performance of rules obtained by using PRIMROSE-REX2: the results in the first row are derived by using both positive and negative rules and those in the second row are derived by only positive rules. The third and fourth row show the results derived by using both positive and negative rules and those by positive rules acquired directly from a medical experts. These results show that the combination of positive and negative rules outperforms positive rules and gains almost the same performance as those experts .

6   Conclusions

In this paper, the characteristics of two measures, classification accuracy and coverage are discussed, which shows that both measures are dual and that accuracy and coverage are measures of both positive and negative rules, respectively.

---

[5] The reason why we select the domain of headache is that we formerly developed an expert system RHINOS (Rule-based Headache INformation Organizing System), which makes a differential diagnosis in headache [14,15]. In this system, it takes about six months to acquire knowledge from domain experts.

Table 4. Evaluation of RH (accuracy: averaged)

Method	Accuracy
PRIMEROSE-REX2 (Positive and Negative)	91.4% (851/930)
PRIMEROSE-REX (Positive)	78.5% (729/930)
RHINOS (Positive and Negative)	93.5% (864/930)
RHINOS (Positive)	82.8% (765/930)

Then, an algorithm for induction of positive and negative rules is introduced. The proposed method was evaluated on medical databases, the experimental results of which show that induced rules correctly represented experts' knowledge and several interesting patterns were discovered.

References

1. Michalski RS, Mozetic I, Hong J, and Lavrac N: The Multi-Purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains. *Proceedings of the fifth National Conference on Artificial Intelligence*, AAAI Press, Palo Alto CA, pp 1041-1045, 1986.
2. Pawlak Z: *Rough Sets*. Kluwer Academic Publishers, Dordrecht, 1991.
3. Quinlan JR: *C4.5 - Programs for Machine Learning*. Morgan Kaufmann, Palo Alto CA, 1993.
4. Tsumoto S and Tanaka H: Automated Discovery of Medical Expert System Rules from Clinical Databases based on Rough Sets. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining 96*, AAAI Press, Palo Alto CA, pp.63-69, 1996.
5. Tsumoto S: Modelling Medical Diagnostic Rules based on Rough Sets. In: Polkowski L and Skowron A (Eds): *Rough Sets and Current Trends in Computing*, Lecture Note in Artificial Intelligence **1424**, 1998.
6. 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.
7. Shavlik JW and Dietterich TG(Eds): *Readings in Machine Learning*. Morgan Kaufmann, Palo Alto CA, 1990.
8. Rissanen J: *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore, 1989.
9. Ziarko W: Variable Precision Rough Set Model. *Journal of Computer and System Sciences* **46**:39-59, 1993.
10. Pawlak Z: Rough Modus Ponens. In: *Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems 98*, Paris, 1998.
11. Adams RD and Victor M: *Principles of Neurology*, 5th edition. McGraw-Hill, New York, 1993.
12. Buchnan BG and Shortliffe EH(Eds): *Rule-Based Expert Systems*. Addison-Wesley, 1984.
13. Tsumoto S: Automated Extraction of Medical Expert System Rules from Clinical Databases based on Rough Set Theory. *Journal of Information Sciences*, **112**, 67-84, 1998.

14. Matsumura Y, Matsunaga T, Hata Y, Kimura M, Matsumura H: Consultation system for diagnoses of headache and facial pain: RHINOS. *Medical Informatics* **11**: 145-157, 1988.
15. Matsumura Y, Matsunaga T, Maeda Y, Tsumoto S, Matsumura H, Kimura M: Consultation System for Diagnosis of Headache and Facial Pain: "RHINOS". *Proceedings of Logic Prgram Conferences*, pp.287-298, 1985.