Automated Extraction of Expert System Rules from Databases based on Rough Set Theory

Shusaku Tsumoto and Hiroshi Tanaka
Department of Information Medicine
Medical Research Institute, Tokyo Medical and Dental University
1-5-45 Yushima, Bunkyo-ku Tokyo 113 Japan
TEL: +81-3-5803-5840 FAX: +81-3-5803-0247
E-mail: tsumoto.com@tmd.ac.jp, tanaka@cim.tmd.ac.jp

Abstract

Automated knowledge acquisition is an important research area to solve the bottleneck problem in developing expert systems. For this purpose, several methods of inductive learning, such as induction of decision trees, AQ method, and neural networks, have been introduced. However, most of the approaches focus on inducing rules which classify cases correctly. On the contrary, medical experts also learn other information which is important for medical diagnostic procedures from databases. In this paper, a rule-induction system, called PRIMEROSE3(Probabilistic Rule Induction Method based on Rough Sets version 3.0), is introduced. This program first analyzes the statistical characteristics of attribute-value pairs from training samples, then determines what kind of diagnosing model can be applied to the training samples. Then, it extracts not only classification rules for differential diagnosis, but also other medical knowledge needed for other diagnostic procedures in a selected diagnosing model. PRIMEROSE3 is evaluated on three kinds of clinical databases and the induced results are compared with domain knowledge acquired from medical experts, including classification rules. The experimental results show that our proposed method correctly not only selects a diagnosing model, but also extracts domain knowledge.

Keywords: learning and data mining, rule induction, rough sets

1. Introduction

One of the most important problems in developing expert systems is knowledge acquisition from experts (Buchanan and Shortliffe, 1984). While there have been developed many knowledge acquisition tools to simplify this process, it is still difficult to automate this process. In order to solve this problem, many inductive learning methods, such as induction of decision trees (Breiman, 1984; Cestnik, et al., 1987; Quinlan, 1986), rule induction methods (Clark and Niblett, 1989; Indurkhya and Weiss, 1991; Michalski, 1983; Michalski, et al. 1986) and rough set theory (Pawlak, 1991; Ziarko, 1991), are introduced in order to discover knowledge from large databases.

However, most of the approaches focus on inducing classification rules, which classifies cases correctly. On the contrary, medical experts also learn other kinds of knowledge, which are important for medical diagnostic procedures, from clinical cases.

In this paper, a rule induction system, called PRIMEROSE3(Probabilistic Rule Induction Method based on Rough Sets version 3.0), is introduced. This program first analyzes the statistical characteristics of attribute-value pairs from training samples, then determines what kind of diagnosing model can be applied to the training samples. Then, it extracts not only classification rules for differential diagnosis, but also other medical knowledge needed for other diagnostic procedures in a selected diagnosing model. PRIMEROSE3 is evaluated on three kinds of clinical databases and the induced results are compared with domain knowledge acquired from medical experts, including classification rules. The experimental results show that our proposed method correctly not only selects a diagnosing model, but also extracts domain knowledge.

The paper is organized as follows: Section 2 discusses rough set theory. Section 3 illustrates three diagnosing models. Section 4 presents our new method, PRIMEROSE3 and Section 5 gives experimental results. Section 6 and Section 7 discuss the problems of PRIMEROSE3 and related work, respectively. Finally, Section 8 concludes this paper.

2. Rough Sets and Rules

2.1 Rough Set Theory

Rough set theory(Pawlak, 1991) clarifies set-theoretic characteristics of the classes over combinatorial patterns of the attributes in order to acquire some sets of attributes for classification and to evaluate how precisely attributes in a database can classify data.

Let us illustrate the main concepts of rough sets which are needed for our formulation. Table 2 is

Table 1: A Small Database

	age	loc	nat	prod	nau	M1	class
1	50-59	occ	per	0	0	1	m.c.h.
2	40-49	who	per	0	0	1	m.c.h.
3	40-49	lat	thr	1	1	0	migra
4	40-40	who	thr	1	1	0	migra
5	40-49	who	rad	0	0	1	m.c.h.
6	50-59	who	per	0	1	1	m.c.h.

DEFINITIONS: loc: location, nat: nature, prod: prodrome, nau: nausea, M1: tenderness of M1, who: whole, occ: occular, lat: lateral, per: persistent, thr: throbbing, rad: radiating, m.c.h.: muscle contraction headache, migra: migraine, 1: Yes, 0: No.

a small database whose patients chiefly complain of headache. First, let us consider how an attribute "loc" classify the headache patients' set of the table. The set whose value of the attribute "loc" is equal to "who"is {2,4,5,6}(In the following, the numbers represent each record number). This set means that we cannot classify {2,4,5,6} further solely by using the constraint $R = [loc = who]$. This set is defined as the indiscernible set over the relation R, denoted by $[x]_R = \{2,4,5,6\}$. In this set, {2,5,6} suffer from muscle contraction headache("m.c.h."), and {4} suffers from migraine("migra"). Hence we need other additional attributes to discriminate between "migra" and "m.c.h." Using this concept, we can evaluate the classification power of each attribute. For example, "prod=1" is specific to the case of migraine ("migra"). We can also extend this indiscernible relation to multivariate cases, such as $[x]_{[loc=who]\wedge[M1=1]} = \{2,5,6\}$ and $[x]_{[loc=who]\vee[M1=1]} = \{1,2,5,6\}$, where $\wedge$ and $\vee$ denote "and" and "or" respectively. In the framework of rough set theory, the set {2,5,6} is called *strictly definable* by the former conjunction, and also called *roughly definable* by the latter disjunctive formula. Therefore, the classification of training samples D can be viewed as a search procedure for the best set $[x]_R$ supported by the relation R. In this way, we can define the characteristics of classification in the set-theoretic framework. For example, accuracy (SI) and coverage, or true positive rate (CI) can be defined as:

$$\alpha_R(D) = \frac{|[x]_R \cap D|}{|[x]_R|}, \text{ and } \kappa_R(D) = \frac{|[x]_R \cap D|}{|D|},$$

where $|D|$ denotes the cardinality of D and where $\alpha_R(D)$ denotes an accuracy of R with respect to classification of D, $SI(R,D)$ and $\kappa_R(D)$ denotes a true positive rate of R to D, $CI(R,D)$, respectively.

For example, let R be equal to $[loc = who]$ and D be equal to a set of "m.c.h." Then, $[x]_R$ and D is equal to $\{2,4,5,6\}$ and $\{1,2,5,6\}$, respectively, and the intersection of both sets, $[x]_R \cap D$, is $\{2,4,5,6\} \cap \{1,2,5,6\} = \{2,5,6\}$. Thus, $\alpha_R(D)$ is equal to $|\{2,5,6\}|/|\{2,4,5,6\}| = 3/4 = 0.75$ and $\kappa_R(D)$ is obtained as: $|\{2,5,6\}|/|\{1,2,5,6\}| = 3/4 = 0.75$, respectively.

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

For further information on rough set theory, readers could refer to (Pawlak, 1991; Ziarko, 1991; Ziarko, 1993).

2.2 Probabilistic Rules

In order to describe diagnosing rules, we first define probabilistic rules, using notations of rough set theory(Pawlak, 1991). To illustrate the main ideas, we use a small database shown in Table 1.

First, a combination of attribute-value pairs, corresponding to a complex in AQ terminology(Michalski, 1983), is denoted by an equivalence relation R_f, which is defined as follows.

Definition 1 (Equivalence Relation) *Let U be a universe, and V be a set of values. A total function f from U to V is called an assignment function of an attribute. Then, we introduce an equivalence relation R_f such that for any $u, v \in U$, $u \equiv R_f v$ iff $f(u) = f(v)$.*

For example, $[age = 50-59]\&[loc = occular]$ will be one equivalence relation, denoted by $R_f = [age = 50-59]\&[loc = occular]$. Secondly, a set of samples which satisfy R_f is denoted by $[x]_{R_f}$, corresponding to a star in AQ terminology. For example, when $\{2,3,4,5\}$ is a set of samples which satisfy $[age = 40-49]$, $[x]_{[age=40-49]}$ is equal to $\{2,3,4,5\}$. [1] Finally, thirdly, U, which stands for "Universe", denotes the whole training samples.

According to this notation, probabilistic rules are defined as follows:

Definition 2 (Probabilistic Rules) *Let R_f be an equivalence relation specified by some assignment function f, D denote a set whose elements belong to a class d, or positive examples in the whole training samples (the universe), U. Finally, let $|D|$ denote the cardinality of D. A probabilistic rule of D is defined as a quadruple, $< R_f \stackrel{\alpha,\kappa}{\rightarrow} d, \alpha_{R_f}(D), \kappa_{R_f}(D) >$, where*

[1] In this notation, "n" denotes the nth sample in a dataset (Table 1).

$R_f \overset{\alpha,\kappa}{\rightarrow} d$ *satisfies the following conditions:*[2]

$$(1) \quad [x]_{R_f} \bigcap D \neq \phi,$$

$$(2) \quad \alpha_{R_f}(D) = \frac{|[x]_{R_f} \bigcap D)|}{|[x]_{R_f}|},$$

$$(3) \quad \kappa_{R_f}(D) = \frac{|[x]_{R_f} \bigcap D)|}{|D|}.$$

In the above definition, α corresponds to the accuracy measure: if α of a rule is equal to 0.9, then the accuracy is also equal to 0.9. On the other hand, κ is a statistical measure of how proportion of D is covered by this rule, that is, a coverage or a true positive rate: when κ is equal to 0.5, half of the members of a class belongs to the set whose members satisfy that equivalence relation.

For example, let us consider a rule $[age = 40-49] \rightarrow m.c.h.$ Since $[x]_{[age=40-49]} = \{2,3,4,5\}$ and $D = \{1,2,5,6\}$, $\alpha_{[age=40-49]}(D) = |\{2,5\}|/|\{2,3,4,5\}| = 0.5$ and $\kappa_{[age=40-49]}(D) = |\{2,5\}|/|\{1,2,5,6\}| = 0.5$. Thus, if a patient, who complains a headache, is 40 to 49 years old, m.c.h. is suspected with accuracy 0.5, and this rule covers 50 % of the cases.

3. Diagnosing Model

3.1 Simplest Diagnosing Model

The simplest diagnosing model is that which only uses classification rules which have high accuracy and high coverage.[3] This model is applicable when rules of high accuracy can be derived. Such rules can be defined as:

$$R \overset{\alpha,\kappa}{\rightarrow} d \quad s.t. \quad R = \vee_i R_i = \vee \wedge_j [a_j = v_k],$$
$$\alpha_{R_i}(D) > \delta_\alpha, \quad and \quad \kappa_{R_i}(D) > \delta_\kappa,$$

where δ_α and δ_κ denote given thresholds for accuracy and coverage, respectively. It is notable that this rule is a kind of probabilistic proposition with two statistical measures, which is one extension of Ziarko's variable precision model(VPRS) (Ziarko, 1993). [4] It is also notable that this model only uses inclusive rules of a RHINOS diagnosing model which is also defined below.

[2]It is notable that this rule is a kind of probabilistic proposition with two statistical measures, which is one kind of an extension of Ziarko's variable precision model(VPRS) (Ziarko, 1993).

[3]In this model, we assume that accuracy is dominant over coverage

[4]In VPRS model, the two precisions of accuracy are given, and the probabilistic proposition with accuracy and two precision conserves the characteristics of the ordinary proposition. Thus, our model is to introduce the probabilistic proposition not only with accuracy, but also with coverage.

3.2 Weak Diagnosing Model

In some probabilistic domains, it is difficult to derive classification rules which have high accuracy. In these cases, the way to overcome this problem is to induce rules which have high coverage. Such rules can be defined as:

$$d \overset{\kappa,\alpha}{\rightarrow} R \quad s.t. \quad R = \vee_i R_i = \vee \wedge_j [a_j = v_k],$$
$$\kappa_{R_i}(D) > \delta_\kappa, \quad and \quad \alpha_{R_i}(D) > \delta_\alpha.$$

Thus, in this model, it is assumed that coverage is dominant over accuracy and this model only induces the necessity condition of diagnosis of class d.

In the subsequent subsection, we introduce the combination of the above two diagnosing models.

3.3 RHINOS Diagnosing Model

RHINOS is an expert system which diagnoses the causes of headache or facial pain from manifestations (Kimura, et al., 1985; Matsumura, et al., 1986). In this system, a diagnosing model proposed by Matsumura is applied, which is composed of the following three kinds of reasoning processes: exclusive reasoning, inclusive reasoning, and reasoning about complications.

Firstly, exclusive reasoning is the one that when a patient does not have a symptom which always appears in any case on a disease, such a disease can be excluded. Secondly, inclusive reasoning is the one that when a patient has symptoms specific to a disease, the disease can be suspected. Finally, thirdly, reasoning about complications is that when some symptoms which cannot be explained by that disease, complications of other diseases can be suspected.

Using the above diagnosing model, we consider three kinds of rules corresponding to each process, which can be described in terms of rough set theory as follows.

(1) Exclusive Rule

$$R \overset{\alpha,\kappa}{\rightarrow} d \quad s.t. \quad R = \wedge_i R_i = \wedge \vee_j [a_j = v_k],$$
$$and \ \kappa_{[a_j=v_k]}(D) = 1.0.$$

Strictly Speaking, this proposition should be written as: $d \rightarrow R$. However, for comparison with other two rules, we choose this notation. In the above example, the relation R of the exclusive rule for "classic" is described as:
$[age = 40-49] \wedge ([loc = lat] \vee [loc = who]) \wedge [nat = thr] \wedge [jolt = 1] \wedge [prod = 1] \wedge [nau = 1] \wedge [M1 = 0]$.

(2) Inclusive Rule

$$R \overset{\alpha,\kappa}{\rightarrow} d \quad s.t. \qquad R = \vee_i R_i = \vee_i \wedge_j [a_j = v_k],$$
$$\alpha_{R_i}(D) > \delta_\alpha, \ and \ \kappa_{R_i}(D) > \delta_\kappa.$$

In the above example, the simplest relation R of the inclusive rule for "classic", is described as: $[nat = thr] \vee [jolt = 1] \vee [M1 = 1]$. However, induction of inclusive rules gives us two problems. First, accuracy and coverage are overfitted to the training samples. Secondly, the above rule is only one of many rules which are induced from the above training samples. Therefore some of them should be selected from primary induced rules under some preference criterion. These problems will be discussed in the next section.

(3) Disease Image:

$$R \overset{\alpha,\kappa}{\rightarrow} d \quad s.t. \qquad R = \vee R_i = \vee [a_i = v_j],$$
$$\alpha_{R_i}(D) > 0 \ (\kappa_{R_i}(D) > 0).$$

In the above example, the relation R of the disease image for "classic" is described as:
$[age = 40 - 49] \vee [loc = lat] \vee [loc = who] \vee [nat = thr] \vee [nau = 1] \vee [jolt = 1] \vee [M1 = 0]$.

It is notable that coverage κ play an important role in the definition of these rules, compared with simplest diagnosing model.

4. PRIMEROSE3

In this section, we introduce a rule-induction system, called PRIMEROSE3(Probabilistic Rule Induction Method based on Rough Sets version 3.0). This program first analyzes the statistical characteristics of attribute-value pairs from training samples, then determines what kind of diagnosing model can be applied to these training samples. Then, it extracts not only classification rules for differential diagnosis, but also other medical knowledge needed for other diagnostic procedures, based on a selected diagnosing model.

4.1 Selection of Diagnosing Model

As discussed in Section 3, coverage plays an important role in selection of a diagnosing model. Thus, PRIMEROSE3 first measures the statistical characteristics of coverage of elementary attribute-value pairs, which corresponds to selectors. Then, it measures the statistical characteristics of accuracy of the whole pattern of attribute-value pairs observed in a dataset.

In this algorithm, we use the following characteristic of coverage.

Table 2: Frequency Table of Coverage

κ	0.00	0.25	0.50	0.75	1.00
m.c.h.	5	6	5	4	2
classic	10	10	6	0	6

Proposition 1 (Monotonicity of Coverage) *Let R_{i+1} denote an attribute-value pair, $R_i \wedge [a_{i+1} = v_j]$. Then,*

$$\kappa_{R_{i+1}}(D) \leq \kappa_{R_i}(D).$$

Proof. Since $[x]_{R_{i+1}} \subseteq [x]_{R_i}$ holds, $\kappa_{R_{i+1}}(D) = \frac{|[x]_{R_{i+1}} \cap D|}{|D|} \leq \frac{|[x]_{R_i} \cap D|}{|D|} = \kappa_{R_i}(D)$. □

Furthermore, in rule induction methods, R_{i+1} is selected to satisfy $\alpha_{R_{i+1}}(D) > \alpha_{R_i}(D)$. Therefore, it is sufficient to check the behavior of coverage of elementary attribute-value pairs in order to estimate which diagnosing model should be selected, while it is necessary to check the behavior of accuracy both of elementary attribute-value pairs and of patterns observed in the databases in order to estimate the characteristics of induced rules. From these considerations, the selection algorithm is defined as follows.

(1) Calculate a coverage and an accuracy of each attribute value pair $[a_i = v_j]$.

(2) Calculate an accuracy of each pattern $\wedge_i [a_i = v_j]$.

(3) Construct a frequency table with respect to coverage and accuracy for each class.

(4) If each class has at least one attribute value pair whose coverage is equal to 1.0 and if more than half of coverage values is larger than δ_κ, goto next. Else, select simplest diagnosing model.

(5) If the median of accuracy of each is larger than δ_α, then select RHINOS diagnosing model and quit. Else select weak diagnosing model and quit.

For the above example shown in Table 1, frequency tables of coverage and accuracy is obtained as Table 2 and Table 3. Thus, let us consider a case when δ_κ is set to 0.5, and δ_α is set to 0.75. Candidates of diagnosing model will be weak diagnosing model and RHINOS diagnosing model from Table 2. Next, from Table 3, RHINOS diagnosing model will be selected, since the median of accuracy of elementary pairs is exactly 1.0.

Table 3: Frequency Table of Accuracy

Elementary Attribute Value Pairs							
α	0.00	0.25	0.33	0.50	0.67	0.75	1.00
m.c.h.	5	0	2	1	4	1	9
classic	9	1	4	1	2	0	5
Patterns Observed in a Dataset							
α	0.00	0.25	0.33	0.50	0.67	0.75	1.00
m.c.h.	0	0	0	0	0	0	4
classic	0	0	0	0	0	0	2

4.2 Algorithm for Rule Induction

For the limitation of space, we only discuss an algorithm for induction of RHINOS rules. However, since the other two models can be viewed as specific forms of RHINOS model, it is easy to derive each algorithm from the induction algorithm shown below.

An induction algorithm for RHINOS rules consists of two procedures. One is an exhaustive search through all the attribute-value pairs (*selectors* in the AQ terminology (Michalski, 1983)), and the other is a heuristic search for inclusive rules through the combinations of all the attribute-value pairs (*complexes* in the AQ terminology).

Exhaustive Search Let D and δ denote training samples of the target class d (*positive examples*) and a threshold to select attributes for inclusive rules. Then, this search procedure is defined as follows.

```
procedure Exhaustive Search;
  var
    L : List; /* A list of elementary relations */
  begin
    L := P0; /* P0: A list of elementary relations */
    while (L ≠ {}) do
      begin
        Select one pair [a_i = v_j] from L;
        if ([x]_[a_i=v_j] ∩ D ≠ φ) then do
          /* D: a set if positive examples */
          begin
            R_di := R_di ∨ [a_i = v_j];
              /* Disease Image */
            if (κ_[a_i=v_j](D) > δ_κ)
              then L_ir := L_ir + {[a_i = v_j]};
              /* Candidates for Inclusive Rules */
            if (κ_[a_i=v_j](D) = 1.0)
              then R_er := R_er ∧ [a_i = v_j];
              /* Exclusive Rule */
          end
        L := L − {[a_i = v_j]};
      end
  end {Exhaustive Search};
```

The above procedure is repeated for all the attribute-value pairs, and computes exclusive rules, disease images, and candidates of inclusive rules. These candidates are input into the heuristic search procedure, discussed in the next subsection.

In the above example in Table 1, let d be "classic", and $[age = 40 - 49]$ be selected as $[a_i = v_j]$. Since $[x]_{[age=40-49]} \cap D(= \{3,4\}) \neq \phi$, this pair is included in the disease image. However, when δ is set to 0.5, this pair is not included in the inclusive rule, because $\alpha_{[age=40-49]}(D) = 0.5$. Finally, since $D \subset [x]_{[age=40-49]}(= \{2,3,4,5\})$, this pair is also included in the exclusive rule.

Next, $[age = 50 - 59]$ is selected. However, this pair will be abandoned since the intersection of $[x]_{[age=50-59]}$ and D is empty, or $[x]_{[age=50-59]} \cap D = \phi$.

Heuristic Search Since the definition of inclusive rules is a little weak, many inclusive rules can be obtained. For the above example, a relation $[nau = 1]$ satisfies $D \cap [x]_{[nau=1]} \neq \phi$, so it is also one of the inclusive rules of "m.c.h.", although accuracy of that rule is equal to 1/3. In order to suppress induction of such rules, which have low classificatory power, only equivalence relations whose accuracy is larger than δ_α is selected. For example, when δ is set to 1/2(=0.5), the above relation $[age = 40 - 49]$ is eliminated from the candidates of inclusive rules, because accuracy of this relation is less than the precision, Furthermore, PRIMEROSE3 minimizes the number of attributes not to include the attributes which do not gain the classificatory power, called *dependent* variables. This procedure can be described as follows:

```
procedure Heuristic Search;
  var
    i : integer;   M, L_i : List;
  begin
    L_1 := L_ir; /* Candidates for Inclusive Rules */
    i := 1;   M := {};
    for i := 1 to n do
        /* n: Total number of attributes */
      begin
        while ( L_i ≠ {} ) do
          begin
            Select one pair R = ∧[a_i = v_j] from L_i;
            L_i := L_i − {R};
            if   (α_R(D) > δ_α)   and   (κ_R(D) > δ_κ)
              then do S_ir := S_ir + {R};
              /* Include R as Inclusive Rule */
            else M := M + {R};
          end
        L_{i+1} := (A list of the whole combination of
                    the conjunction formulae in M);
      end
  end {Heuristic Search};
```

In the above example in Table 1, the coverage of $[M1 = 1]$ for "m.c.h" is maximum. Furthermore, since $\alpha_{[M1=1]}(D) = 1.0$, it is included in inclusive rules of "m.c.h". The next maximum one is $[nau = 0]$, whose coverage is equal to 3/4. Since this accuracy is also equal to 1.0, it is also included in inclusive rules. At this point, we have two inclusive rules as follows:

$[M1 = 1] \overset{\alpha=1.0,\kappa=1.0}{\rightarrow}$ "*m.c.h.*" and $[nau = 0] \overset{\alpha=1.0,\kappa=0.75}{\rightarrow}$ "*m.c.h.*" Repeating these procedures, all the inclusive rules are acquired.

4.3 Estimation of Accuracy and Coverage

The above definition of accuracy and coverage shows that small training samples cause the overestimates of accuracy and coverage. In the above example shown in Table 1, both of accuracy and coverage of the simplest rule for "classic" are equal to 1.0. This means that this rule correctly diagnoses and covers all the cases of the disease "classic". However, in general, these meanings hold only in the world of the small training samples. In this sense, accuracy and coverage are biased. Thus, these biases should be corrected by introducing other estimating methods, since the biases cannot be detected by the induced method.

Note that this problem is similar to that of error rates of discriminant function in multivariate analysis (Mclachlan, 1992), the field in which resampling methods are reported to be useful for the estimation.

Hence the resampling methods are applied to estimation of accuracy and coverage, as shown in the following subsection.

4.4 Cross-Validation and the Bootstrap method

Cross-validation method for error estimation is performed as following: first, the who training samples $\mathcal{L}$ are split into V blocks: $\{\mathcal{L}_1, \mathcal{L}_2, \cdots, \mathcal{L}_V\}$. Secondly, repeat for V times the procedure in which rules are induced from the training samples $\mathcal{L} - \mathcal{L}_i (i = 1, \cdots, V)$ and examine the error rate err_i of the rules using $\mathcal{L}_i$ as test samples. Finally, the whole error rate err is derived by averaging err_i over i, that is, $err = \sum_{i=1}^{V} err_i/V$ (this method is called V-fold cross-validation). Therefore this method for estimation of accuracy or coverage can be used by replacing the calculation of err by that of accuracy or coverage, respectively, and by regarding test samples as unobserved cases.

On the other hand, the Bootstrap method is executed as follows: first, empirical probabilistic distribution(F_n) is generated from the original training samples (Efron, 1982). Secondly, the Monte-Carlo method is applied and training samples are randomly taken by using F_n. Thirdly, rules are induced by using new training samples. Finally, these results are tested by the original training samples and then statistical measures, such as error rate are calculated. These four steps are iterated for finite times. Empirically, it is shown that about 200 times repetition is sufficient for estimation (Efron, 1982).

Table 4: Information about Databases

Domain	Samples	Classes	Attributes
headache	1477	10	20
meningitis	198	3	25
CVD	261	6	27

Interestingly, Efron shows that estimators by 2-fold cross-validation are asymptotically equal to predictive estimators for completely new pattern of data, and that Bootstrap estimators are asymptotically equal to maximum likelihood estimators and are a little overfitted to training samples (Efron, 1983). Hence, the former estimators can be used as the lower bound of accuracy and coverage, and the latter as the upper bound of accuracy and coverage.

Furthermore, in order to reduce the high variance of estimators by cross-validation, we introduce repeated cross-validation method, which is first introduced by Walker (Walker and Olshen, 1992). In this method, cross-validation methods are executed repeatedly(safely, 100 times), and estimates are averaged over all the trials. In summary, since our strategy is to avoid the overestimation and the high variabilities, combination of repeated 2-fold cross-validation and the Bootstrap method is adopted in this paper.

5. Experimental Results

PRIMEROSE3 is applied to headache(RHINOS's domain), meningitis, and cerebrovascular diseases, whose precise information is given in Table 4. The experiments are performed by the following three procedures. First, these samples are randomly split into pseudo-training samples and pseudo-test samples. Secondly, by using the pseudo-training samples, PRIMEROSE3 selects a model, induces rules and the statistical measures.[5] Thirdly, the induced results are tested by the pseudo-test samples. These procedures are repeated for 100 times and average each accuracy and the estimators for accuracy of diagnosis over 100 trials.

Table 5 shows the results about selection of a diagnosing model. In the domain of headache and meningitis, the dominant model is RHINOS diagnosing model and the simplest model, respectively, which matches with decision of medical experts. On the other hand, both models can be applied in the domain of CVD, although experts select the RHINOS diagnosing model. In the subsequent subsections, we only discuss the re-

[5] The thresholds δ_α and δ_κ is set to 0.75 and 0.5, respectively in these experiments.

Table 5: Selection of Diagnosing Model

Domain	Simplest	Weak	RHINOS	Experts' Model
headache	11	7	82	RHINOS
meningitis	65	13	22	Simplest
CVD	43	18	39	RHINOS

Table 6: Experimental Results on Headache (Averaged)

Method	ER	IR	DI	CPU time
PR3	95.0%	88.3%	93.2%	10.9 min
RHINOS	98.0%	95.0%	97.4%	–
C4.5	–	85.8%	–	15.6 min
CN2	–	87.0%	–	17.3 min
AQ15	–	86.2%	–	17.2 min

DEFINITIONS: PR3: PRIMEROSE3
ER: Exclusive Rule Accuracy
IR: Inclusive Rule Accuracy
DI: Disease Image Accuracy

sults of two domains in which PRIMEROSE3 correctly selects a diagnosing model.

5.1 Headache

Experimental results on headache are summarized in Table 6 and 7. In Table 6, the first column, exclusive rule accuracy denotes how many training samples that do not belong to a class are excluded correctly from the candidates. The second column is equivalent to the averaged classification accuracy and the third column shows how many symptoms, which cannot be explained by diagnostic conclusions, are detected by the disease image. And the fourth column shows CPU time needed to induced rules. The first row of Table 6 is the result of PRIMROSE3, and the second one is that of medical experts. And, for comparison, we compare the classification accuracy of inclusive rules with that of C4.5 (Quinlan, 1993), CN2 (Clark and Niblett, 1989) and AQ-15 (Michalski, et al. 1986), which is shown in the third to fifth row of Table 6. Table 7 shows the results of estimation derived by using repeated cross-validation method (R-CV) and the bootstrap method (BS).

These results is summarized to the following four points. First, the induced inclusive rules perform worse than those of medical experts, exclusive rules and disease images gain the same performance, compared with experts' rules. Secondly, our method performs a little better than four classical empirical learning methods, although the differences are not statistically significant.

Table 7: Experimental Results of Estimation

Method	Accuracy	R-CV	BS
PR3	88.3%	78.7%	91.6%
C4.5	85.8%	77.2%	90.7%
CN2	87.0%	72.9%	93.2%
AQ15	86.2%	74.6%	92.3%

Table 8: Experimental Results of Meningitis (Averaged)

Method	Accuracy	CPU time
PRIMEROSE3	83.9%	69 sec
Expert Rules	93.0%	–
C4.5	74.0%	57 sec
CN2	75.0%	60 sec
AQ15	84.7%	62 sec

Thirdly, PRIMEROSE3 is faster than the other methods. Finally, fourthly, R-CV estimator and BS estimator can be regarded as the lower boundary and the upper boundary of each rule accuracy. Hence the interval of these two estimators can be used as the estimator of performance of each rule.

5.2 Meningitis

Table 8 and 9 show the experimental results derived from databases on meningitis. The first row of Table 8 is the result of PRIMROSE3, and the second one is that of medical experts. And, for comparison, we compare the classification accuracy of inclusive rules with that of C4.5 (Quinlan, 1993), CN2 (Clark and Niblett, 1989) and AQ-15 (Michalski, et al. 1986), which is shown in the third to fifth row.

In Table 9, the results of estimation are derived by using repeated cross-validation method (R-CV) and the bootstrap method (BS).

As shown in Table 8, PRIMEROSE 3 performs as well as the other methods.

5.3 CVD

Experimental results on CVD are summarized in Table 10 and 11. The notations of both tables are the same as those of Table 6 and Table 7.

These results is summarized to the following four points. First, the induced inclusive rules perform worse than those of medical experts, exclusive rules and disease images gain the same performance, compared with experts' rules. Secondly, our method performs a little better than four classical empirical learning methods,

Table 9: Experimental Results of Estimation

Method	Accuracy	R-CV	BS
PR3	88.3%	78.7%	91.6%
C4.5	74.0%	70.9%	85.7%
CN2	75.0%	71.2%	85.7%
AQ15	84.7%	75.6%	87.3%

Table 10: Experimental Results on CVD (Averaged)

Method	ER	IR	DI	CPU time
PR3	91.0%	84.3%	94.3%	59 sec
RHINOS	97.5%	92.9%	93.6%	–
C4.5	–	79.7%	–	91 sec
CN2	–	78.7%	–	72 sec
AQ15	–	78.9%	–	81 sec

DEFINITIONS: PR3: PRIMEROSE3
ER: Exclusive Rule Accuracy
IR: Inclusive Rule Accuracy
DI: Disease Image Accuracy

although the differences are not statistically significant. Thirdly, PRIMEROSE3 is faster than the other methods. Finally, fourthly, R-CV estimator and BS estimator can be regarded as the lower boundary and the upper boundary of each rule accuracy. Hence the interval of these two estimators can be used as the estimator of performance of each rule.

6. Discussion

6.1 Exclusive Rule

As discussed in Section 3, we intend to formulate induction of exclusive rules by using the whole given attributes, although the original exclusive rules are described by the six basic questions, as shown in Appendix. Therefore induced exclusive rules have the maximum number of attributes whose conjunction R also satisfies $\kappa_R(D) = 1.0$. If this maximum combination includes the six basic attributes as a subset, then this selection of basic attributes is one of good choices of attributes, although redundant. Otherwise, the given six attributes may be redundant or the induced results may be insufficient. For the above example shown in Table 1, the maximum combination of attributes is {age, loc, nat, jolt, prod, nau, M1 } is included in both exclusive rules.

On the contrary, in the database for the above experiments, the maximum combination is 13 attributes, derived as follows: Age, Pain location, Nature of the pain, Severity of the pain, History since onset, Existence of jolt headache, Tendency of depression, and Tenderness of M1 to M6, which is a superset of the six basic attributes. Thus, this selection can be a good choice.

Table 11: Experimental Results of Estimation

Method	Accuracy	R-CV	BS
PR3	84.3%	68.7%	88.6%
C4.5	79.7%	67.2%	87.7%
CN2	78.7%	62.3%	89.2%
AQ15	78.9%	64.9%	92.3%

In this way, the induction of maximum combination can be also used as a "rough" check of induced results or our diagnosing model on exclusive rules, which can be formulated as below. [6]

Let A and E denote a set of the induced attributes for exclusive rules and a set of attributes acquired from domain experts. Thus, the following four relations can be considered. First, if $A \subset E$, then A is insufficient or E is redundant. Secondly, if $A = E$, then both sets are sufficient to represent diagnosing model in an applied domain. Thirdly, if $A \supset E$, then A is redundant or E is insufficient. Finally, fourthly, if intersection of A and E is not empty ($A \cap E \neq \phi$), then either or both sets are insufficient.

Reader may say that the above relations are weak and indeterminate. However, the above indefinite parts should be constrained by information on domain knowledge. For example, let us consider the case when $A \subset E$. When E is validated by experts, A is insufficient in the first relation. However, in general, E can be viewed as A obtained by large samples, and $A \supset E$ should hold, which shows that a given database is problematic. Moreover, the constraint on exclusive rules, $\kappa_R(D) = 1.0$, suggests that there exist a class which does not appear in the database, because the already given classes cannot support $\kappa_R(D) = 1.0$, that is, $[x]_R \cap D \neq D$ will hold in the future.

On the other hand, when E is not well given by experts and A is induced from sufficiently large samples, E will be redundant, which means that the proposed model for E does not fit to this database or this domain.

This kind of knowledge is important, because we sometimes need to know whether samples are enough to induce knowledge and whether an applied inducing model is useful to analyze databases.

[6] This discussion assumes that the whole attributes are sufficient to classify the present and the future cases into given classes.

Thus, the above four relations give simple examinations to check the characteristics of samples and the applicability of a given diagnosing model. It is our future work to develop more precise checking methodology for automated knowledge acquisition.

6.2 Precision for Inclusive Rules

In the above experiments, a threshold δ for selection of inclusive rules is set to 0.5 because of the following reason. Since the database on headache supports 10 classes, the naive "*a priori*" probability for each class is equal to $1/10 = 0.1$. Thus, when the probability of one disease should be set to 0.5, the other disease is suspected with probability $0.5/9 \approx 0.055$, which means that the assigned accuracy will become half.

Although this precision contributes to the reduction of computational complexity, this methodology, which gives a threshold in a static way, cause a serious problem. For example, there exists a case when the accuracy for the first, the second, and the third candidate is 0.5, 0.49, and 0.01, whereas accuracy for other classes is almost equal to 0. Formally, provided an attribute-value pair, R, the following equations hold: $\alpha_R(D_1) = 0.5, \alpha_R(D_2) = 0.49, \alpha_R(D_3) = 0.01$, and $\alpha_R(D_i) \approx 0 (i = 4, \cdots, 10)$. Then, both of the first and the second candidate should be suspected because those accuracies are very close, compared with the accuracy for the third and other classes. However, if a threshold is statically set to 0.5, then this pair is not included in positive rules for D_2. In this way, a threshold should be determined dynamically for each attribute-value pair. In the above example, an attribute-value pair should be included in positive rules of D_1 and D_2.

From discussion with domain experts, it is found that this type of reasoning is very natural, which may contribute to the differences between induced rules and ones acquired from medical experts. Thus, even in a learning algorithm, comparison between the whole given classes should be included in order to realize more plausible reasoning strategy.

Unfortunately, since the proposed algorithm runs for each disease independently, the above type of reasoning cannot be incorporated in a natural manner, which causes computational complexity to be higher. It is also our future work to develop such interacting process in the learning algorithm.

7. Related Work

7.1 AQ Algorithm

AQ is an inductive learning system based on incremental STAR algorithm(Michalski, 1983). This algorithm selects one "seed" from positive examples and starts from one "selector"(attribute-value pair) contained in this "seed" example. It adds selectors incrementally until the "complexes" (conjunction of attributes) explain only positive examples, called a **bounded star**. Since many complexes can satisfy these positive examples, AQ finds the most preferred ones, according to a flexible extra-logical criterion.

It would be worth noting that the positive examples which supports the complexes corresponds to the lower approximation, or the positive region in rough set theory. That is, the rules induced by AQ are equivalent to consistent rules defined by Pawlak when neither constructive generalization(Michalski, 1983; Wnek and Michalski) nor truncation(Michalski, et al. 1986) are used, and when the length of STAR is not restricted. As a matter of fact, AQ's star algorithm without constructive generalization can be reformulated by the concepts of rough sets. For example, a bounded star denoted by $G(e|U - D, m_0)$ in Michalski's notation is equal to $G = \{R_i | [x]_{R_i} = D_j\}$, such that $|G| = m_0$ where $|G|$ denotes the cardinality of G. This star is composed of many complexes, which is ordered by LEF_i, lexicographic evaluation functional, which is defined as the following pair:$< (-negcov, \tau_1), (poscov, \tau_2) >$ where *negcov* and *poscov* are numbers of negative and positive examples, respectively, covered by an expression in the star, and where τ_1 and τ_2 are tolerance threshold for criterion $poscov, negcov$ ($\tau \in [0..100\%]$). This algorithm shows that AQ method is a kind of greedy algorithm which finds independent variables using selectors which are equivalent to equivalence relations in terms of rough sets.

Thus, our heuristic search method is very similar to AQ method, while our method uses statistical measures, rather than LEF criterion, which implicitly includes the notions of accuracy and coverage. The difference between our heuristic search procedure and AQ method is that PRIMEROSE3 explicitly uses accuracy and coverage and that it only uses elementary attribute-value pairs selected by the exhaustive search procedure, according to the characteristics of coverage, although AQ implicitly uses the criteria for both measures. The main reason why PRIMEROSE3 uses statistical measures is that discussion about the statistical characteristics of both measures is easier and that the definition of probabilistic rules is much clearer. As shown in Section 4, three kinds of rules are easily classified into three category with respect to accuracy and coverage. Especially, since coverage plays an important role in the classification of rules, it is very easy to implement an induction algorithm of exclusive rules and disease image. Thus, PRIMEROSE3 can be viewed as a combination of AQ algorithm and the exhaustive

search method.

7.2 Discovery of Association Rules

Mannila et al.(Mannila, 1994) report a new algorithm for discovery of association rules, which is one class of regularities, introduced by Agrawal et al.(Agrawal, et. al, 1993). Their method is very similar to ours with respect to the following two points.

(1) Association Rules The concept of association rules is similar to our induced rules. Actually, association rules can be described in the rough set framework.

That is, we say that an association rule over r (training samples) satisfies $W \Rightarrow B$ with respect to γ and σ, if

$$|[x]_W \cap [x]_B| \geq \sigma n, \quad (1)$$

and

$$\frac{|[x]_W \cap [x]_B|}{|[x]_W|} \geq \gamma, \quad (2)$$

where n, γ, and σ denotes the size of training samples, confidence threshold, and support threshold, respectively. Also, W and B denotes an equivalence relation and a class, respectively. Furthermore, we also say that W is *covering*, if

$$|[x]_W| \geq \sigma n. \quad (3)$$

It is notable that the left side of the above formulae (6) and (8) correspond to the formula (3) as to κ, coverage, and the left side of the formula (7) corresponds to (2) as to α, accuracy. The only difference is that we classify rules, corresponding to association rules, into three categories: exclusive rules, inclusive rules, and disease image.

The reason why we classify these rules is that this classification reflects the diagnosing model of medical experts, by which the computational speed of diagnostic reasoning is higher.

(2) Mannila's Algorithm Mannila et al. introduce an algorithm to find association rules based on Agrawal's algorithm. The main points of their algorithms are database pass and candidate generation. Database pass produces a set of attributes L_s as the collection of all covering sets of size s in C_s. Then, candidate generation calculates C_{s+1}, which denotes the collection of all the sets of attributes of size s, from L_s. Then, again, database pass is repeated to produce L_{s+1}. The effectiveness of this algorithm is guaranteed by the fact that all subsets of a covering set are covering.

The main difference between Mannila's algorithm and PRIMEROSE3 is that Mannila uses the check algorithm for covering to obtain association rules, whereas we use both accuracy and coverage to compute and classify rules.

In the discovery of association rules, all of the combination of attribute-value pairs in C_s have the property of covering. On the other hand, our algorithm does not focus on the above property of covering. It removes an attribute-value pair which has both high accuracy and high coverage. That is, PRIMEROSE3 does not search for regularities which satisfy covering, but search for regularities important for classification.

Thus, interestingly enough, when many attribute-value pairs have the covering property, or covers many training samples, Mannila's algorithm will be slow, although PRIMEROSE3 algorithm will be fast in this case. When few pairs cover many training samples, Mannila's algorithm will be fast, and our system will not be faster.

8. Conclusion

In this paper, we introduce a system, called PRIMEROSE3 which selects a diagnosing model, extracts not only classification rules for differential diagnosis, but also other medical knowledge which is needed for diagnosis, based on the selected diagnosing model. We evaluate this system by using three clinical databases, and compared the induced results with rules acquired from medical experts. The results show that our proposed method correctly induce RHINOS rules and estimate the statistical measures of rules.

Acknowledgements

The authors would like to thank the reviewers for their insightful comments. This research is supported by Grants-in-Aid for Scientific Research No.06680343 from the Ministry of Education, Science and Culture in Japan.

References

Agrawal, R., Imielinski, T., and Swami, A. Mining association rules between sets of items in large databases, *Proceedings of the 1993 International Conference on Management of Data (SIGMOD 93)*, pp. 207-216, 1993.

Breiman, L., Freidman, J., Olshen, R., and Stone, C. (1984). *Classification And Regression Trees.* Belmont, CA: Wadsworth International Group.

Buchanan, B. G. and Shortliffe, E. H.(eds.) (1984). *Rule-Based Expert Systems*, Addison-Wesley.

Cestnik, B., Kononenko, I., Bratko, I. (1987). Assistant 86: A knowledge elicitation tool for sophisticated

users. *Proceedings of the Second European Working Session on Learning*, pp.31-45, Sigma Press.

Clark, P., Niblett, T. (1989). The CN2 Induction Algorithm. *Machine Learning*, **3**,261-283.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics, Pennsylvania.

Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross validation. *Journal of American Statistics Association*, **78**, 316-331.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule ? *Journal of American Statistics Association*, **82**, 171-200.

Kimura, M., et al. (1985). RHINOS: A Consultation System for Diagnoses of Headache and Facial Pain: RHINOS. *Proceedings of the ninth International Joint Conference on Artificial Intelligence*, 393-396, Morgan Kaufmann, CA.

Indurkhya, N. and Weiss, S. (1991). Iterative Rule Induction Methods, *Applied Intelligence*, **1**, 43-54.

Mannila, H., Toivonen, H., Verkamo, A.I. Efficient Algorithms for Discovering Association Rules, *Proceedings of Knowledge Discovery in Databases (KDD-94)*, pp.181-192, AAAI press, CA, 1994.

Matsumura, Y., et al. (1986). Consultation system for diagnoses of headache and facial pain: RHINOS. *Medical Informatics*, **11**, 145-157.

Mclachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition.* John Wiley and Sons, New York.

Michalski, R. S. (1983). A Theory and Methodology of Machine Learning. Michalski, R.S., Carbonell, J.G. and Mitchell, T.M., *Machine Learning - An Artificial Intelligence Approach.* Morgan Kaufmann, Palo Alto, CA.

Michalski, R. S., Mozetic, I., Hong, J., and Lavrac, N. (1986). The Multi-Purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains. *Proceedings of the fifth National Conference on Artificial Intelligence*, 1041-1045, AAAI Press, Palo Alto, CA.

Pawlak, Z. (1991). *Rough Sets.* Kluwer Academic Publishers, Dordrecht.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, **1**, 81-106.

Quinlan, J.R. (1993). *C4.5 - Programs for Machine Learning*, Morgan Kaufmann, CA.

Walker, M. G. and Olshen, R. A. (1992). Probability Estimation for Biomedical Classification Problems. *Proceedings of the sixteenth Symposium on Computer Applications on Medical Care*, McGrawHill, New York.

Wnek, J. and Michalski, R.Z. Hypothesis-Driven Constructive Induction in AQ17-HCI: A Method and Experiments, *Machine Learning*, **14**, 139-168, 1994.

Ziarko, W. (1991). The Discovery, Analysis, and Representation of Data Dependencies in Databases, in: Shapiro, G. P. and Frawley, W. J.(eds), *Knowledge Discovery in Databases*, AAAI press, Palo Alto, CA, pp.195-209.

Ziarko, W. (1993). Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, **46**, 39-59.

Appendix: RHINOS Rules

Using the diagnosing model introduced by (Matsumura, et al., 1986), we consider the following three kinds of rules corresponding to each process and develop the following algorithms for acquiring these rules from medical experts.

(1)Exclusive Rule This rule corresponds to exclusive reasoning. In other words, the premise of this rule is equivalent to the necessity condition of the diagnostic conclusion. From the discussion with medical experts, we selected the following six basic attributes which are minimally indispensable to defining the necessity condition: *1. Age, 2. Pain location, 3. Nature of the pain, 4. Severity of the pain, 5. History since onset, 6. Existence of jolt headache.*

For example, the exclusive rule of common migraine is the following:

```
In order to suspect common migraine,
the following symptoms are required:
pain location: not eyes and
nature: throbbing or persistent or radiating
and history: paroxysmal or sudden and
jolt headache: positive.
```

One of the reason why we selected the six attributes is to solve the interface problem of expert systems: if the whole attributes are considered, we also have to input all the symptoms which are not needed for diagnosis. To make exclusive reasoning compact, the only

minimal requirements are chosen. It is notable that this kind of selection can be viewed as the ordering of given attributes, and it is expected that such ordering can be induced from databases. Therefore, in PRIMEROSE3, an algorithm for induction of exclusive rules scans the whole given attributes. It is because the minimal requirements for describing exclusive rules can be computed after all the exclusive rules are induced. Furthermore, this ordering can be viewed as a "rough" check of induced results and applicability of our diagnosing model. This issue is discussed in Section 6.

(2)Inclusive Rule This rule consists of a set of positive rules, the premises of which are composed of a set of manifestations specific to a disease to be included. If a patient satisfy one set, we suspect this disease with some probability. This rule is derived from medical experts by using the following algorithm for each disease: *1. Take a set of manifestations by which we strongly suspect a disease. 2. Set the probability that a patient has the disease with this set of manifestations:SI(Satisfactory Index) 3. Set the ratio of the patients who satisfy the set to all the patients of this disease:CI(Covering Index) 4. If sum of the derived CI(tCI) is equal to 1.0 then end. If not, goto 5. 5. For the patients of this disease who do not satisfy all the collected set of manifestations, goto 1.* Therefore a positive rule is described by a set of manifestations, its satisfactory index (SI), which corresponds to *accuracy measure*, and its covering index (CI), which corresponds to *total positive rate* or *coverage*. Note that SI and CI are given empirically by medical experts.

Formally, each positive rule is represented as a quadruple:
$\langle d, R_i, SI_i, CI_i \rangle$, where d denotes its conclusion, and R_i denotes its premise. Hence each inclusive rule is described as: $\langle \{\langle d, R_1, SI_1(, CI_1)\rangle, \cdots, \langle d, R_k, SI_k(, CI_k)\rangle\}, tCI\rangle$, where total CI(tCI) is defined as the CI of a total rule, composed of a disjunctive formula of all rules, $R_1 \vee R_2 \vee \cdots \vee R_k$.

Let us show the inclusive rule of common migraine (tCI=0.9) as an example, which is composed of the following three rules:

```
If history: paroxysmal, jolt headache:
positive, nature: throbbing or persistent,
prodrome: no, intermittent symptom: no,
persistent time: more than 6 hours, and
pain location: not eye,
then common migraine is suspected
with accuracy 0.9 (SI=0.9) and this rule
covers 60 percent of the total cases
(CI=0.6).

If history: paroxysmal, jolt headache:
positive, nature: throbbing or persistent,
prodrome: no, intermittent symptom: no,
and pain location: not eye,
then common migraine is suspected
with accuracy 0.8 (SI=0.8) and this rule
covers 80 percent of the total cases
(CI=0.8).

If history: sudden, jolt headache:
positive, nature: throbbing or persisted,
and prodrome: no,
then common migraine is suspected
with accuracy 0.5 (SI=0.5) and this rule
covers 30 percent of the total case
(CI=0.3).
```

In the above rules, tCI shows that the disjunctive form of above three rules covers 90 percent of total cases of common migraine.[7]

It also means that 10 percent of common migraine cannot be diagnosed by the above rules.

(3)Disease Image This rule is used to detect complications of multiple diseases, acquired by all the possible manifestations of a disease. Using this rule, we search for the manifestations which cannot be explained by the diagnosed disease. Those symptoms suggests complications of other diseases. For example, the disease image of common migraine is shown as follows:

```
The following symptoms can be explained
by common migraine:
pain location: any or
tendency of depression : negative or
jolt headache: positive or ........
```

Therefore, when a patient who suffers from common migraine is depressing, it is suspected that he or she may also have other disease, because this symptom cannot be explained by common migraine.

As shown above, these algorithms are straightforward, and is based on set-theoretic framework.

[7]Since tCI is based on total coverage by the disjunctive form of rules, it is not equal to total sum of CI values of all rules.