Automated Discovery of Medical Expert System Rules from Clinical Databases based on Rough Sets

Shusaku Tsumoto and Hiroshi Tanaka
Department of Information Medicine, Medical Research Institute,
Tokyo Medical and Dental University,
1-5-45 Yushima, Bunkyo-city Tokyo 113 Japan.
E-mail: tsumoto.com@mri.tmd.ac.jp, tanaka@cim.tmd.ac.jp

Abstract

Automated knowledge acquisition is an important research issue to solve the bottleneck problem in developing expert systems. Although many inductive learning methods have been proposed for this purpose, most of the approaches focus only on inducing classification rules. However, medical experts also learn other information important for diagnosis from clinical cases. In this paper, a rule induction method is introduced, which extracts not only classification rules but also other medical knowledge needed for diagnosis. This system is evaluated on a clinical database of headache, whose experimental results show that our proposed method correctly induces diagnostic rules and estimates the statistical measures of rules.

Introduction

One of the most important problems in developing expert systems is knowledge acquisition from experts(Buchanan and Shortliffe 1984). In order to automate this problem, many inductive learning methods, such as induction of decision trees(Breiman, et al. 1984; Quinlan 1993), rule induction methods(Michalski 1983; Michalski, et al. 1986; Quinlan 1993) and rough set theory(Pawlak 1991; Ziarko 1993), are introduced and applied to extract knowledge from databases, which shows that these methods are appropriate.

However, most of the approaches focus only on inducing classification rules, although medical experts also learn other information important for medical diagnostic procedures. Focusing on their learning procedures, Matsumura et al. propose a diagnostic model, which consists of three reasoning processes, and develop an expert system, called RHINOS(Rule-based Headache and facial pain INformation Organizing System)(Matsumura, et al. 1986).

Since RHINOS diagnostic processes are found to be based on the concepts of set theory, it is expected that a set-theoretic approach can describe this diagnostic model and knowledge acquisition procedures.

In order to characterize these procedures, we introduce the concepts of rough set theory, which clarifies set-theoretic characteristics of the classes over combinatorial patterns of the attributes, precisely discussed by (Pawlak 1991). Based on this theory, we develop a program, called PRIMEROSE-REX (Probabilistic Rule Induction Method based on Rough Sets and Resampling methods for Expert systems), which extracts rules for an expert system from clinical databases, and applies resampling methods to the estimation certainty factors of derived rules.[1]

This system is evaluated on the datasets of RHINOS domain. The results show that the proposed method induces RHINOS diagnostic rules correctly from databases and that resampling methods can estimate the performance of these rules and certainty factors.

The paper is organized as follows: in Section 2, we discuss RHINOS diagnostic model. Section 3 shows rough set theory and representation of RHINOS rules based on this theory. Section 4 presents an algorithm for induction of RHINOS diagnostic rules. Section 5 gives experimental results. Section 6 and Section 7 discusses the problems of our work and related work, respectively. Finally, Section 8 concludes this paper.

RHINOS

RHINOS is an expert system which diagnoses clinical cases on headache or facial pain from manifestations. In this system, a diagnostic model proposed by Matsumura(Matsumura, et al. 1986) consists of the following three kinds of reasoning processes: exclusive reasoning, inclusive reasoning, and reasoning about complications.

First, exclusive reasoning excludes a disease from candidates when a patient does not have a symptom which is necessary to diagnose. Secondly, inclusive reasoning suspects a disease in the output of the exclusive process when a patient has symptoms specific to a disease. Finally, reasoning about complications suspects complications of other diseases when some symptoms which cannot be explained by the diagnostic conclusion are obtained.

[1]This system is an extension of PRIMEROSE, which induces classification rules from databases, based on rough sets and resampling methods(Tsumoto and Tanaka 1995).

Each reasoning is rule-based, and all the rules needed for the diagnostic processes are acquired from medical experts in the following way.

(1)Exclusive Rules The premise of an exclusive rule is equivalent to the necessity condition of a diagnostic conclusion. From the discussion with medical experts, we select 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 defined as:

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

One of the reason why we select the six attributes is to solve the interface problem of expert systems: if the whole attributes are considered, we also have to input 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, which can be induced from databases automatically. Therefore we intend to formulate induction of exclusive rules by using the whole given attributes. After the induction, the minimal requirements for describing exclusive rules can be acquired.

(2)Inclusive Rules The premises of inclusive rules are composed of a set of manifestations specific to a disease to be included. If a patient satisfies one set of symptoms, we suspect this disease with some probability. This rule is derived by asking the following items for each disease to the medical experts: *1. a set of manifestations by which we strongly suspect a disease. 2. the probability that a patient has the disease with this set of manifestations:SI(Satisfactory Index) 3. the ratio of the patients who satisfy the set to all the patients of this disease:CI(Covering Index) 4. If the total sum of the derived CI(tCI) is equal to 1.0 then end. Otherwise, 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*. Note that SI and CI are given empirically by medical experts.

For example, one of three positive rules for common migraine is given as follows.

```
If history: paroxysmal, jolt headache: yes,
nature: throbbing or persistent,
prodrome: no, intermittent symptom: no,
```

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.

```
persistent time: more than 6 hours,
and 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).
```

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

```
The following symptoms can be explained by
common migraine: pain location: any or
depressing: not or jolt headache: yes or ...
```

Therefore, when a patient who suffers from common migraine is depressing, it is suspected that he or she may also have other disease.

As shown above, three kinds of rules are straightforward, and an inducing algorithm is expected to be implemented on computers easily. Thus, we introduce rough set theory in order to describe these algorithms as shown in the next section.

Formalization of Rules

Probabilistic Rules

In order to describe three kinds of diagnostic rules, we first define probabilistic rules, using the following three 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, which is corresponding to a complex in AQ (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 $u, v \in U$, uR_fv 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 $\{2,3,4,5\}$. [2]

Finally, thirdly, U, which stands for "Universe", denotes all training samples.

According to these notations, 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 all 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 $R_f \stackrel{\alpha,\kappa}{\rightarrow} d$ satisfies the following conditions:* [3]

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

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

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

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 belong to the set whose members satisfy that equivalence relation.

For example, let us consider a case of a proposition $[age = 40-49] \rightarrow m.c.h.$ Since $[x]_{[age=40-49]}$ is equal to $\{2,3,4,5\}$ and D is equal to $\{1,2,5,6\}$, $\alpha_{[age=40-49]}(D) = |\{2,5\}|/|\{2,3,4,5\}| = 0.5$ 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, then m.c.h. is suspected, whose accuracy and coverage are equal to 0.5.

RHINOS Diagnostic Rules

By the use of these notations, RHINOS diagnostic rules are described in the following way.

(1) Exclusive rules: $R \stackrel{\alpha,\kappa}{\rightarrow} d$ *s.t.* $R = \wedge_i R_i = \wedge \vee_j [a_j = v_k]$, *and* $\kappa_{R_i}(D) = 1.0$. [4] In the above example, the relation R for migraine is described as: $[age = 40-49] \wedge ([location = lateral] \vee [location = whole]) \wedge [nature = throbbing] \wedge ([history = paroxysmal] \vee [history = persistent]) \wedge [jolt = yes] \wedge [prod = yes] \wedge [nau = yes] \wedge [M1 = no] \wedge [M2 = no]$.

(2) Inclusive rules: $R \stackrel{\alpha,\kappa}{\rightarrow} d$ *s.t.* $R = \vee_i R_i = \vee \wedge_j \vee_k [a_j = v_k]$, $\alpha_{R_i}(D) > \delta_\alpha$, *and* $\kappa_{R_i}(D) > \delta_\kappa$.

In the above example, the simplest relation R for migraine, is described as: $[nature = throbbing] \vee [history = paroxysmal] \vee [jolt = yes] \vee [M1 = yes]$. However, induction of inclusive rules gives us two problems. First, SI and CI 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: $R \stackrel{\alpha,\kappa}{\rightarrow} d$ *s.t.* $R = \vee R_i \vee [a_i = v_j]$, *and* $\alpha_{R_i}(D) > 0$ ($\kappa_{R_i}(D) > 0$).

In the above example, the relation R for migraine is described as:
$[age = 40-49] \vee [location = lateral] \vee [location = whole] \vee [nature = throbbing] \vee [severity = strong] \vee [severity = weak] \vee [history = paroxysmal] \vee [nausea = yes] \vee [jolt = yes] \vee [M1 = no] \vee [M2 = no]$.

As shown in the formal definition of these rules, a coverage $\kappa_R(D)$ play an important role in classification of diagnostic rules.

Induction of Rules

An induction algorithm of RHINOS rules consists of two procedures. One is an exhaustive search procedure to induce the exclusive rule and the disease image for each disease through all the attribute-value pairs, corresponding to *selectors* in AQ (Michalski 1983), and the other is a postprocessing procedure to induce inclusive rules through the combinations of all the attribute-value pairs, which corresponds to *complexes* in AQ.

Exhaustive Search

Let D denote training samples of the target class d, or *positive examples*. This search procedure is defined as shown in Figure 1. In the above example in Table 1, let d be migraine and $[age = 40-49]$ be selected as $[a_i = v_j]$. Since the intersection $[x]_{[age=40-49]} \cap D(=$

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

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

[4] Strictly Speaking, this proposition should be written as: $d \rightarrow R$. However, for comparison with other two rules, we choose this notation.

```
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};
```

Figure 1: An Algorithm for Exhaustive Search

```
procedure Postprocessing Procedure;
  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)
              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 {Postprocessing Procedure };
```

Figure 2: An Algorithm for Postprocessing Procedure

$\{3,4\}$) is not equal to ϕ, this pair is included in the disease image. However, since $\alpha_{[age=40-49]}(D) = 0.5$, this pair is not included in the inclusive rule. Finally, since $D \subset [x]_{[age=40-49]} (= \{2,3,4,5\})$, this pair is also included in the exclusive rule.

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

When all the attribute-value pairs are examined, not only the exclusive rule and disease image shown in the above section, but also the candidates of inclusive rules are also derived. The latter ones are used as inputs of the second procedure.

Postprocessing Procedure

Because the definition of inclusive rules is a little weak, many inclusive rules can be obtained. In the above example, an equivalence 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 SI 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 SI is larger than 0.5 are selected. For example, since the above relation $[age = 40-49]$ is less than this precision, it is eliminated from the candidates of inclusive rules. Furthermore, PRIMEROSE-REX 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 shown in Figure 2. In the above example in Table 1, the coverage of an attribute-value pair $[prod = 0]$ for "m.c.h" takes a maximum value. Furthermore, since the accuracy $\alpha_{[prod=0]}(D)$ is equal to 1.0, it is included in inclusive rules of "m.c.h". The next maximum one is $[M1 = 1]$, whose coverage is equal to 1.0. 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: $[prod = 0] \stackrel{\alpha=1.0,\kappa=1.0}{\rightarrow}$ *"m.c.h."* and $[M1 = 1] \stackrel{\alpha=1.0,\kappa=1.0}{\rightarrow}$ *"m.c.h."* Repeating these procedures, all the inclusive rules are acquired.

Estimation of Statistical Measures

The above definition of statistical measures shows that small training samples causes their overestimation. In the above example, both of the measures are equal to 1.0. This means that this rule correctly diagnoses and covers all the cases of the migrane. However, in general, these meanings hold only in the world of the small training samples. In this sense, accuracy and coverage are biased. Thus, we should correct these biases 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 (Efron 1982), 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.

Cross-Validation and the Bootstrap

Cross-validation method for error estimation is performed as following: first, all 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 coverage and accuracy can be used by replacing the calculation of err by that of coverage and accuracy, 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 statistical measures, such as error rate are calculated. These four steps are iterated for finite times. Empirically, it is shown that repeating these steps for 200 times is sufficient for estimation (Efron 1982).

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 1982). Hence, the former estimators can be used as the lower bounds of both measures, and the latter as their upper bounds.

Furthermore, in order to reduce the high variance of estimators by cross validation, we introduce repeated cross validation method, which is firstly introduced by Walker (Walker and Olshen 1992). In this method, cross validation methods are executed repeatedly (safely, 100 times)(Tsumoto and Tanaka 1995), 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.

Experimental Results

We apply PRIMEROSE-REX to the following three medical domains: headache(RHINOS domain), whose training samples consist of 1477 samples, 10 classes, and 20 attributes, cerebulovasular diseases, whose training samples consist of 620 samples, 15 classes, and 25 attributes, and meningitis, whose training samples consists of 213 samples, 3 classes, and 27 attributes. In these experiments, δ_α and δ_κ are set to 0.75 and 0.5, respectively. The experiments are performed by the following four procedures. First, these samples are randomly split into half (new training samples) and half (new test samples). For example, 1477 samples are split into 738 training samples and 739 training samples. Secondly, PRIMEROSE-REX, AQ15 and CART are applied to the new training samples. Thirdly, the repeated cross validation method and the bootstrap method are applied to the new training samples in order to estimate the accuracy and coverage of PRIMEROSE-REX. Finally, the induced results are tested by the new test samples. These procedures are repeated for 100 times and all the estimators are averaged over 100 trials.

Table 2: Experimental Results (Headache)

Method	ER-A	IR-A	DI-A
PR-REX	95.0%	88.3%	93.2%
Experts	98.0%	95.0%	97.4%
CART	–	85.8%	–
AQ15	–	86.2%	–
R-CV	72.9%	78.7%	83.8%
BS	98.4%	91.6%	95.6%

DEFINITIONS: PR-REX: PRIMEROSE-REX,
ER-A: Exclusive Rule Accuracy,
IR-A: Inclusive Rule Accuracy,
DI-A: Disease Image Accuracy

Table 3: Experimental Results (Cerebulovasculuar Diseases)

Method	ER-A	IR-A	DI-A
PR-REX	91.0%	84.3%	94.3%
Experts	97.5%	92.9%	93.6%
CART	–	79.7%	–
AQ15	–	78.9%	–
R-CV	72.9%	78.7%	83.8%
BS	93.4%	92.5%	95.9%

Experimental results are shown in Table 2 to 4. Exclusive rule accuracy(ER-A) means how many training samples that do not belong to a class are excluded correctly from the candidates. Inclusive rule accuracy(IR-A) is equivalent to the averaged classification accuracy. Finally, disease image accuracy(DI-A) shows how many symptoms, which cannot be explained by diagnostic conclusions, are detected by the disease image. The first row is the results obtained by using PRIMROSE-REX, and the second one is the results derived from medical experts. And, for comparison, we compare the classification accuracy of inclusive rules with that of CART and AQ-15, which is shown in the third and fourth row. Finally, in the fifth and sixth row, we present the results of estimation by repeated cross-validation method (R-CV) and the bootstrap method (BS). These results can be summarized to the following three points. First, the induced rules perform a little worse than those of medical experts. Secondly, our method performs a little better than classical empirical learning methods, CART and AQ15. Finally, thirdly, R-CV estimator and BS estimator can be regarded as the lower boundary and

Table 4: Experimental Results (Meningitis)

Method	ER-A	IR-A	DI-A
PR-REX	88.9%	82.5%	92.6%
Experts	95.4%	93.2%	96.7%
CART	–	81.4%	–
AQ15	–	82.5%	–
R-CV	64.3%	61.3%	73.8%
BS	89.5%	93.2%	98.2%

the upper boundary of each rule accuracy. Hence the interval of these two estimators can be used as the estimators of accuracy and coverage of each rule.

Discussion

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. 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, location, nature, history, jolt, prod, nau, M1, M2 }. [5] Since this set does not include an attribute "severity", the six given attributes or the induced results are insufficient in this small database. In this case, however, the sixth attributes are acquired by medical experts through a large number of experienced cases. Thus, the induced attributes should be revised by using additional samples in the future.

On the contrary, in the database on headache, 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.

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 in the following way. [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. Second, if $A = E$, then both sets are sufficient to represent a diagnosing model in an applied domain. Third, if $A \supset E$, then A is redundant or E is insufficient. Finally, fourth, 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.

Thus, the above four relations give a simple examination 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.

Related Work

Discovery of Association Rules

Mannila et al.(Mannila, et al. 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 our method with respect to the use of set-theoretical operations.

(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, a confidence threshold, and a support threshold,

[5] Severity cannot be a member, since $[sever = weak] \vee [sever = strong]$ is included in both exclusive rules.

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

respectively. Also, W and B denote 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 diagnostic model of medical experts, which makes the computational speed of diagnostic reasoning higher.

(2) Mannila's Algorithm: Mannila introduces an algorithm to find association rules based on Agrawal's algorithm (Mannila, et al. 1994). The main points of their algorithm are the following two procedures: 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, the candidate generation calculates C_{s+1}, which denotes the collection of all the sets of attributes of size s, from L_s. Then, again, the database pass procedure 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 PRIMEROSE-REX 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 the combinations 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 selects an attribute-value pair which has both high accuracy and high coverage. That is, PRIMEROSE-REX does not search for regularities which satisfy covering, but search for regularities important for classification.

Thus, interestingly, when many attribute-value pairs have the covering property, or covers many training samples, Mannila's algorithm will be slow, although PRIMEROSE-REX 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 slower.

Acknowledgements

This research is supported by Grants-in-Aid for Scientific Research No.08680388 from the Ministry of Education, Science and Culture in Japan.

References

Agrawal, R., Imielinski, T., and Swami, A. (1993). 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.

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.

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

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

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

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.

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.

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

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

Tsumoto, S. and Tanaka, H.(1994). Induction of Medical Expert System Rules based on Rough Sets and Resampling Methods. *Proceedings of the 18th Symposium on Computer Applications on Medical Care*(Washington, D.C.), pp.1066-1070. Philadelphia: Hanley & Belfus, INC., November.

Tsumoto, S. and Tanaka, H. (1995). PRIMEROSE: Probabilistic Rule Induction Method based on Rough Sets and Resampling Methods. *Computational Intelligence*, **11**, 389-405.

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.

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