Discovery of Knowledge about Drug Side Effects in Clinical Databases based on Rough Set Model

Shusaku Tsumoto
Department of Information Medicine, Medical Research Institute,
Tokyo Medical and Dental University
1-5-45 Yushima, Bunkyo-ku Tokyo 113 Japan
E-mail: tsumoto@computer.org

Abstract

Rule discovery methods have been introduced to find useful and unexpected patters from databases. However, one of the most important problems on these methods is that extracted rules have only positive knowledge, which do not include negative information that medical experts need to confirm whether a patient will suffer from symptoms caused by drug side-effect. This paper first discusses the characteristics of medical reasoning and defines positive and negative rules based on rough set model. Then, algorithms for induction of positive and negative rules are introduced. Then, the proposed method was evaluated on clinical databases, the experimental results of which shows several interesting patterns were discovered, such as a rule describing a relation between urticaria caused by antibiotics and food.

Introduction

Rule induction methods are classified into two categories, induction of deterministic rules and probabilistic ones(Michalski 1986; Pawlak 1991; Tsumoto and Tanaka 1996). While deterministic rules are supported by positive examples, probabilistic ones are supported by large positive examples and small negative samples. That is, both kinds of rules select positively one decision if a case satisfies their conditional parts.

However, domain experts do not use only positive reasoning but also negative reasoning, since a domain is not always deterministic. For example, when a patient does not have a headache, migraine should not be suspected: negative reasoning plays an important role in cutting the search space of a differential diagnosis(Tsumoto and Tanaka 1996).[1] Therefore, negative rules should be induced from databases in order to induce rules which will be easier for domain experts to interpret: induction of plausible rules will be important for an interaction between domain experts and rule induction methods.

In this paper, first, the characteristics of medical reasoning are focused and two kinds of rules, positive rules and negative rules, are defined as a model of medical reasoning. Both rules, whose supporting sets correspond to the lower and upper approximation in rough sets(Pawlak 1991), are defined as deterministic rules with two measures, classification accuracy and coverage. Then, algorithms for induction of positive and negative rules are introduced, which are defined as search procedures using accuracy and coverage as evaluation functions. 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.

Focusing Mechanism

One of the characteristics in medical reasoning is a focusing mechanism, which is used to select the final diagnosis from many candidates(Tsumoto and Tanaka 1996). 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 modeled as usage of two kinds of rules, negative rules (exclusive rules) and positive rules, the former of

[1] The essential point is that if extracted patterns do not reflect experts' reasoning process, domain experts have difficulties in interpreting them. Without interpretation of domain experts, a discovery procedure would not proceed, which also means that the interaction between human experts and computers is indispensable to computer-assisted discovery.

[2] Relations this diagnostic model with another diagnostic model are discussed in (Tsumoto 1998).

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.

Probabilistic Rules

Accuracy and Coverage

In the subsequent sections, we adopt the following notations, which is introduced in (Skowron 1994).

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, $A = (U, A \cup \{d\})$.

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

By the use of this framework, 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 $R \rightarrow d$ is defined as:

$$\begin{aligned} \alpha_R(D) &= \frac{|R_A \cap D|}{|R_A|}(= P(D|R)), \text{ and} \\ \kappa_R(D) &= \frac{|R_A \cap D|}{|D|}(= P(R|D)), \end{aligned}$$

where $|A|$ denotes the cardinality of a set A, $\alpha_R(D)$ denotes a classification accuracy of R as to classification of D, and $\kappa_R(D)$ denotes a coverage, or a true positive rate of R to D, respectively.

It is notable that these two measures are equal to conditional probabilities: accuarcy is a proability of D under the condition of R, coverage is one of R under the condition of D. It is also 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.[3]

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 $R \leftrightarrow D$.

Also, Pawlak recently reports a Bayesian relation between accuracy and coverage(Pawlak 1998b):

$$\begin{aligned} \alpha_R(D)P(D) &= P(R|D)P(D) = P(R, D) \\ &= P(R)P(D|R) = \kappa_R(D)P(R) \end{aligned}$$

This relation also suggests that *a priori* and *a posteori* probablities should be easily and automatically calculated from databases. By the use of accuracy and coverage, a probabilistic rule is defined as:

$$R \overset{\alpha,\kappa}{\rightarrow} d \quad s.t. \quad R = \wedge_j \vee_k [a_j = v_k],\ \alpha_R(D) \geq \delta_\alpha,\ \kappa_R(D) \geq \delta_\kappa.$$

This rule is a kind of probabilistic proposition with two statistical measures, which is an extension of Ziarko's variable precision model(VPRS) (Ziarko 1993).[4]

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.

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(Pawlak 1991). Thus, a positive rule is represented as:

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

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.

Negative Rules

Before defining a negative rule, let us first introduce an exclusive rule, the contrapositive of a negative rule(Tsumoto and Tanaka 1996). An exclusive rule is defined as a rule supported by all the positive examples, the coverage of which is equal to 1.0.[5] It is notable that the set supporting a exclusive rule corresponds to the upper approximation of a target concept, which is introduced in rough sets(Pawlak 1991). Thus, an exclusive rule is represented as:

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

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

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

[3]These characteristics are from formal definiton of accuracy and coverage. In this paper, these measures are important not only from the viewpoint of propositional logic, but also from that of modelling medical experts' reasoning, as shown later.

[4]This probabilistic rule is also a kind of *Rough Modus Ponens*(Pawlak 1998).

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

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. In summary, a negative rule is defined as:

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

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(Pawlak 1991).

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(Tsumoto and Tanaka 1996), as shown in Figure 1. Negative rules are derived as the contrapositive of induced exclusive rules. On the other hand, positive rules are induced as inclusive rules by the algorithm introduced in PRIMEROSE-REX, as shown in Figure 2. For induction of positive rules, the threshold of accuracy and coverage is set to 1.0 and 0.0, respectively.

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 $(\kappa_{[a_i=v_j]}(D) > 0)$ **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 $[a_i = v_j]$ in Exclusive Rule */
          end
      $L := L - [a_i = v_j]$;
      end
    Construct Negative Rules:
        Take the contrapositive of R_{er}.
  end {*Exclusive and Negative Rules*};

Figure 1: Induction of Exclusive and Negative Rules

Experimental Results

For experimental evaluation, a new system, called PRIMEROSE-REX2 (Probabilistic Rule Induction Method for Rules of Expert System ver 2.0), is developed, where the algorithms discussed in Section 4 are implemented. PRIMEROSE-REX2 was applied to the following two datasets on drug side-effects: allergy for antibiotics, whose training samples consist of 31119 samples, 4 classes and 137 attributes, side-effects of steroid, whose training samples consist of 3620 samples, 11 classes and 285 attributes.

What is Discovered ?

Positive Rules in Antibiotics In the domain of allergic reactions to antibiotics, the following positive rules, which medical experts do not expect, are obtained.

$$[Sex = F] \wedge [Food = Fish] \rightarrow [Effect = Urticaria]$$
$$[Age \leq 40] \wedge [Food = Fish] \rightarrow [Effect = Urticaria]$$

The most interesting points are that the above two rules have information about age and sex, which often seems to be unimportant attributes for allergic reactions.

The first discovery is that women do often suffer from urticaria compared with men, since such relationships between sex and meningitis has not been discussed in medical context(Adams and Victor 1993). The second discovery is that $[age < 40]$ is also an important factor to suspect allergic reactions.

These results were also re-evaluted in medical practice. Recently, the above two rules were checked by additional 121 cases who suffered from drug allergy to antibiotics. Surprisingly, the above rules misclassfied only 21 cases that is, the total accuracy is equal to $100/121 = 82.6\%$. The validation of these rules is still ongoing, which will be reported in the near future.

Negative Rules in Steroid Dataset Concerning the database on side-effects of steroid, several interesting negative rules are derived. The most interesting results are the following positive and negative rules for thalamus hemorrahge:

$$\neg[Sex = Female] \wedge \neg[OGTT = DM]$$
$$\rightarrow \neg[Effect = DM]$$

$$\neg[Hypertension = yes] \wedge \neg[Sex = Female]$$
$$\rightarrow \neg[Effect = HT]$$

Interestingly, $[OGTT = DMpattern]$ under the condition of $[Sex = Female]$ is an important factor to diagnose Steroid induced Diabetus Mellitus. In this do-

```
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 ≠ {} ) 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 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};
```

Figure 2: Induction of Positive Rules

main, 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 will be behind these rules.

Discussion

Positive rules are exactly equivalent to a deterministic rules, which are defined in (Pawlak 1991). So, the disjunction of positive rules corresponds to the positive region of a target concept (decision attribute). On the other hand, negative rules correspond to the negative region of a target concept. From this viewpoint, probabilistic rules correspond to the combination of the boundary region and the positive region (mainly the boundary region).

Thus our approach, the combination of positive and negative deterministic rules captures the target concept as the combination of positive and negative information. Interestingly, our experiment shows that the combination outperforms the usage of only positive rules, which suggests that we need also negative information to achieve higher accuracy. So, although our method is very simple, it captures the important aspect of experts' reasoning and points out that we should examine the role of negative information in experts' decision more closely.

Another aspect of experts' reasoning is fuzzy or probabilistic: in the rough set community, the problems of deterministic rules are pointed by Ziarko(Ziarko 1993), who introduces Variable Precision Rough Set Model (VPRS model). VPRS model extends the positive concept with the precision of classification accuracy: a relation, the classification accuracy of which is larger than a given precision (threshold), will be regarded as positive. Thus, in this model, rules of high accuracy are included in an extended positive region. Analogously, we can also extend the negative concept with the precision of coverage, which will make an extended negative region. The combination of those positive and negative rules will extend the approach introduced in this paper, which is expected to gain the performance or to extract knowledge about experts' decision more correctly. Thus, it will be a future work to check whether the combination of extended positive and negative rules will outperform that of positive and negative deterministic rules.

References

Adams RD. and Victor M. 1993. *Principles of Neurology*, 5th edition, New York, McGraw-Hill.

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.

Pawlak, Z. 1998. Rough Modus Ponens. *Proceedings of IPMU'98* , Paris.

Pawlak, Z. 1998b. Rough Sets and Decision Analysis, *Fifth IIASA workshop on Decision Analysis and Support*, Laxenburg.

Shavlik, J. W. and Dietterich, T.G. (eds.) *Readings in Machine Learning*, CA: Morgan Kaufmann, 1990.

Skowron, A. and Grzymala-Busse, J. 1994. 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.

Tsumoto, S. and Tanaka, H. 1996. Automated Discovery of Medical Expert System Rules from Clinical Databases based on Rough Sets. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining 96, pp.63-69, AAAI Press.

Tsumoto, S. 1998. Modelling Medical Diagnostic Rules based on Rough Sets, *Rough Sets and Current Trends in Computing*, Lecture Note in Artificial Intelligence.

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