

Discovery of Approximate Medical Knowledge 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. One of the most important problems on rule induction methods is that extracted rules do not plausibly represent information on experts' decision processes, which makes rule interpretation by domain experts difficult. This paper first discusses the characteristics of medical reasoning and defines positive and negative rules which models medical experts' rules. Then, algorithms for induction of positive and negative rules are introduced. The proposed method was evaluated on medical databases, the experimental results of which show that induced rules correctly represented experts' knowledge and several interesting patterns were discovered.

1 Introduction

Rule induction methods are classified into two categories, induction of deterministic rules and probabilistic ones[2, 3, 5, 7]. 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[7].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

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.

upper approximation in rough sets[3], 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.

2 Focusing Mechanism

One of the characteristics in medical reasoning is a focusing mechanism, which is used to select the final diagnosis from many candidates[7]. 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 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.

3 Definition of Rules

3.1 Rough Sets

In the following sections, we use the following notations of rough set theory[3], which is illustrated by a small database shown in Table 1, collecting the patients who complained of headache. First, a combination of attribute-value pairs, corresponding to a complex in AQ terminology[6], is denoted by a formula R . For example, $[location = whole] \wedge [nausea = no]$ will be one formula, denoted by $R = [location = whole] \wedge [nausea = no]$.

Secondly, a set of samples which satisfy R is denoted by $[x]_R$, corresponding to a star in AQ terminology. For example, the set, $[x]_{[location=whole]}$, each member of which satisfies $[location = whole]$, is equal to $\{2,4,5,6\}$, which shows that the second, fourth, fifth and sixth case (In the following, the numbers in a set are used to represent each record number). These relations can be also extended to multivariate cases, such as such as $[x]_{[location=whole] \wedge [nausea=no]} = \{2, 5\}$ and

² Relations this diagnostic model with another diagnostic model are discussed in [8].

Table 1. An Example of Database

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

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

$[x]_{[location=whole] \vee [nausea=no]} = \{1, 2, 4, 5, 6\}$, where \wedge and \vee denote “and” and “or” respectively. Finally, U , which stands for “Universe”, denotes all training samples. In the framework of rough set theory, the set $\{2,5\}$ 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 for the best set $[x]_R$ which is supported by the relation R . In this way, we can define the characteristics of classification in the set-theoretic framework.

For further information on rough set theory, readers could refer to [3, 9].

3.2 Classification Accuracy and Coverage

Classification accuracy and coverage (true positive rate) are defined as:

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

where $|A|$, $\alpha_R(D)$ and $\kappa_R(D)$ denote the cardinality of a set A , a classification accuracy of R as to classification of D and a coverage, or a true positive rate of R to D , respectively. In the above example, when R and D are set to $[nau = 1]$ and $[class = migraine]$, $\alpha_R(D) = 2/3 = 0.67$ and $\kappa_R(D) = 2/2 = 1.0$.

It is notable that $\alpha_R(D)$ measures the degree of the sufficiency of a proposition, $R \rightarrow D$, 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 $R \leftrightarrow D$.

3.3 Probabilistic Rules

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

$$R \xrightarrow{\alpha, \kappa} d \quad \text{s.t.} \quad R = \bigwedge_j [a_j = v_k], \alpha_R(D) \geq \delta_\alpha \text{ and } \kappa_R(D) \geq \delta_\kappa,$$

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

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

3.4 Positive Rules

A positive rule is defined as a rule supported by only positive examples, the classification accuracy of which is equal to 1.0. It is notable that the set supporting this rule corresponds to a subset of the lower approximation of a target concept, which is introduced in rough sets[3]. Thus, a positive rule is represented as:

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

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

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

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

3.5 Negative Rules

Before defining a negative rule, let us first introduce an exclusive rule, the contrapositive of a negative rule[7]. An exclusive rule is defined as a rule supported by all the positive examples, the coverage of which is equal to 1.0.⁴ It is notable that the set supporting an exclusive rule corresponds to the upper approximation of a target concept, which is introduced in rough sets[3]. Thus, an exclusive rule is represented as:

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

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

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

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

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

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

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

³ This probabilistic rule is also a kind of *Rough Modus Ponens*[4].

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

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

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

In summary, a negative rule is defined as:

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

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

Negative rules should be also included in a category of deterministic rules, since their coverage, a measure of negative concepts is equal to 1.0. It is also notable that the set supporting a negative rule corresponds to a subset of negative region, which is introduced in rough sets[3].

4 Algorithms for Rule Induction

The contrapositive of a negative rule, an exclusive rule is induced as an exclusive rule by the modification of the algorithm introduced in PRIMEROSE-REX[7], 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[7], 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.

5 Experimental Results

5.1 Performance of Rules Obtained

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 three medical domains: headache(RHINOS domain), whose training samples consist of 1477 samples, 10 classes and 20 attributes, cerebulo-vascular 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.

The experiments were performed by the following three procedures. First, these samples were randomly splits into new training samples and new test samples. Second, using the new training samples, PRIMEROSE-REX2 induced positive and negative rules. Third, the induced results were tested by the new test samples. These procedures were repeated for 100 times and average all the estimators over 100 trials.

Experimental results are shown in Table 2. The first and second row show the results obtained by using PRIMROSE-REX2: the results in the first row

```

procedure Exclusive and Negative Rules;
  var
    L : List; /* A list of elementary attribute-value pairs */
  begin
    L := Pb; /* Pb: A list of elementary attribute-value pairs given in a database */
    while (L ≠ { }) do
      begin
        Select one pair [ai = vj] from L;
        if ( $(x)_{[a_i=v_j]} \cap D \neq \phi$ ) then do /* D: positive examples of a target class d */
          begin
            Lir := Lir + [ai = vj]; /* Candidates for Positive Rules */
            if ( $\kappa_{[a_i=v_j]}(D) = 1.0$ )
              then Rer := Rer ∧ [ai = vj]; /* Include [ai = vj] in a list of Exclusive Rule */
            end
          L := L - [ai = vj];
          end
        Construct Negative Rules:
        Take the contrapositive of Rer.
      end {Exclusive and Negative Rules};

```

Fig. 1. Induction of Exclusive and Negative Rules

Table 2. Experimental Results (Accuracy: Averaged)

Method	Headache	CVD	Meningitis
PRIMEROSE-REX2 (Positive+Negative)	91.3%	89.3%	92.5%
PRIMEROSE-REX2 (Positive)	68.3%	71.3%	74.5%
Experts	95.0%	92.9%	93.2%

were derived by using both positive and negative rules and those in the second row were derived by only positive rules. The third row shows the results derived from medical experts. These results show that the combination of positive and negative rules outperforms positive rules, although it is a little worse than medical experts' rules.

6 What is Discovered ?

6.1 Positive Rules in Meningitis

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

$$\begin{aligned}
 & [WBC < 12000] \& [Sex = Female] \& [Age < 40] \& [CSF_CELL < 1000] \rightarrow Virus \\
 & [Age \geq 40] \& [WBC \geq 8000] \& [Sex = Male] \& [CSF_CELL \geq 1000] \rightarrow Bacteria
 \end{aligned}$$

```

procedure Positive Rules;
  var
    i : integer;  M, Li : List;
  begin
    L1 := Lir;
    /* Lir: 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 ( Li ≠ {} ) do
          begin
            Select one pair  $R = \wedge[a_i = v_j]$  from Li;
            Li := Li - {R};
            if ( $\alpha_R(D) > \delta_\alpha$ )
              then do Sir := Sir + {R}; /* Include R in a list of the Positive Rules */
              else M := M + {R};
            end
            Li+1 := (A list of the whole combination of the conjunction formulae in M);
          end
        end {Positive Rules};
      end

```

Fig. 2. Induction of Positive Rules

The most interesting points are that these rules have information about age and sex, which often seems to be unimportant attributes for differential diagnosis.

The first discovery is that women do not often suffer from bacterial infection, compared with men, since such relationships between sex and meningitis has not been discussed in medical context[1]. Examined the database of meningitis closely, it is found that most of the above patients suffer from chronic diseases, such as DM, LC, and sinusitis, which are the risk factors of bacterial meningitis. The second discovery is that [*age* < 40] is also an important factor not to suspect viral meningitis, which also matches the fact that most old people suffer from chronic diseases.

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

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

6.2 Positive and Negative Rules in CVD

Concerning the database on CVD, several interesting rules are derived. The most interesting results are the following positive and negative rules for thalamus hemorrhage:

$$\begin{aligned} [Sex = Female] \& [Hemiparesis = Left] \& [LOC : positive] \rightarrow Thalamus \\ \neg [Risk : Hypertension] \& \neg [Sensory = no] \rightarrow \neg Thalamus \end{aligned}$$

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

7 Discussion

As discussed in Section 4, positive (PR) and negative rules (NR) are:

$$\begin{aligned} PR : \bigwedge_j [a_j = v_k] \rightarrow d \quad s.t. \quad \alpha_{\bigwedge_j [a_j = v_k]}(D) = 1.0 \\ NR : \bigwedge_j \neg [a_j = v_k] \rightarrow \neg d \quad s.t. \quad \forall [a_j = v_k] \quad \kappa_{[a_j = v_k]}(D) = 1.0. \end{aligned}$$

Positive rules are exactly equivalent to a deterministic rules, which are defined in [3]. 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[9], 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.

Another interest is a measure of boundary region: a measure of positive information is accuracy and one of negative information is coverage. Probabilistic rules can be measured by the combination of accuracy and coverage [7, 8], but the combination of two measures is difficult to compare each rule: to measure the quality of boundary. It will also be one of the important future research directions.

8 Conclusions

In this paper, the characteristics of two measures, classification accuracy and coverage is discussed, which shows that both measures are dual and that accuracy and coverage are measures of both positive and negative rules, respectively. Then, an algorithm for induction of positive and negative rules is introduced. The proposed method was evaluated on medical databases, the experimental results of which show that induced rules correctly represented experts' knowledge and several interesting patterns were discovered.

References

1. Adams RD. and Victor M. *Principles of Neurology*, 5th edition, New York, McGraw-Hill. 1993.
2. Michalski, R. S., Mozetic, I., Hong, J., and Lavrac, N. The Multi-Purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains. *Proceedings of the fifth National Conference on Artificial Intelligence*, 1041-1045, AAAI Press, Palo Alto, CA, 1986.
3. Pawlak, Z. *Rough Sets*. Kluwer Academic Publishers, Dordrecht, 1991.
4. Pawlak, Z. Rough Modus Ponens. *Proceedings of IPMU'98*, Paris, 1998.
5. Quinlan, J.R. *C4.5 - Programs for Machine Learning*, Morgan Kaufmann, CA, 1993.
6. Shavlik, J. W. and Dietterich, T.G. (eds.) *Readings in Machine Learning*, CA: Morgan Kaufmann, 1990.
7. Tsumoto, S. and Tanaka, H. 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, 1996.
8. Tsumoto, S. Modelling Medical Diagnostic Rules based on Rough Sets, *Rough Sets and Current Trends in Computing*, Lecture Note in Artificial Intelligence, 1998.
9. Ziarko, W. Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, **46**, 39-59, 1993.