Rule Extraction using GA-based Fuzzy Modeling

HUGANG HAN[†], YOSHIO MORIOKA[†] and KAZUMI TAKANO[‡]
[†] School of Business, Hiroshima Prefectural University
562 Nanatsuka-cho, Shobara-shi, Hiroshima 727-0023, JAPAN

[‡] Faculty of Health Sciences, Hiroshima Prefecutral College of Health Sciences
1-1 Gakuen-cho, Mihara-shi, Hiroshima 723-0053, JAPAN

Abstract: - The concept of rough set, which an upper/lower approximation are involved in, is giving a powerful tool to extract rules from a database or examples. In order to determine the upper/lower approximation of a subset in an approximation space, each discrete attribute value (usually it appeals like digital figure) plays a very important role in the process of approximation. However, so far few papers pay attention to how to convert an originally continuous attribute values into some appropriate discrete values. Clearly, the better conversion we give, the better approximation accuracy we will get. To do so, we introduce fuzzy logic to divide the continuous attribute values, and further, genetic algorithm (GA) is adopted to obtain the most proper fuzzy division. In this paper, the detailed GA-based fuzzy modeling approach to extract rules from a database is given.

Key-Words: - Rough set; Information system; Database; Rule extraction; GA; Fuzzy modeling

1 Introduction

The effective use of computers in various realms of human activities strongly depends on the efficiency of algorithms implemented in these computers. So far, many theoretical foundation stones for the algorithm have been set up in which the fuzzy set theory is the most attractive one [1]. In a couple of decades, the application of fuzzy set to the field of engineering has obtained some interesting and useful results [2]-[5]. Compared with the performance obtained from the field of engineering, the application of fuzzy set theory to the intelligence systems appears not as good as we were expected. On the other hand, based on the concept of rough classification, the rough set theory was proposed by Pawlak [6]-[7]. Although the formation of rough set is around 20-year later than fuzzy set, its applications to intelligence system such as rule extraction are appearing with great force [8]-[16]. In the most views, the theories of rough set and fuzzy set are related but distinct and complementary [17]. In this paper, we are not interested in the connections and differences of the both theories, but want to pay attention to extract rules from database fusing their merits. Basically, we use rough set theory, which takes into consideration the indiscernibility between objects. The indiscernibility is typically characterized by an equivalence relation, which is obtained by the discrete values of each attribute. And based on the equivalence relation, a subset in the approximation space is approximated by an upper approximation and a lower approximation. Then we can induce some certain rules from the lower approximation, and possible rules from upper approximation. Obviously, the discrete values of each attribute influence the rule extraction directly. There is not any problem when the attribute values are originally discrete. For example, consider an attribute, say "sex", the set of attribute values is $\{0, 1\}$ where 0 expresses "male" and 1 expresses "female". However, how can we get an appropriate set of an attribute if the attribute values are originally continuous? For example, say "age", the candidates of the sets are $\{0, 1\}$, $\{0, 1, 2\}$ and so on where 0 expresses "young", 1 expresses "old", and 2 expresses "middle". Which one should be taken? Even we can choose one, then, what interval age is "middle"? You say 30 $\sim$ 45-year old, and I may argue that is 28 $\sim$ 39-year old. Who is correct? We do not have a clue but experts do. Actually there is not always an expert! Here in this paper we recall the merits of the fuzzy set theory, which deals with the ill-definition of the boundary through the so-called fuzzy membership functions. Further, genetic algorithm (GA) is adopted to obtain the most proper fuzzy division.

The remainder of this paper is arranged as follows. Section 2 describes the regular way to show how to extract rules based on the concepts of rough set and information system. In Section 3, after describing the problem to be considered, we make GA-based fuzzy modeling clear to meet the goal of rule extraction with a proposal of problem solution, and conclusion is given in Section 4.

2 Rule Extraction

2.1 Rough set

An approximation space A is an order pair $A = (U, R)$ where U is a set called universe and R is an equivalence relation over U. Equivalence relation R is also called an indiscernibility relation. If $(x, y) \in U$ and $(x, y) \in R$ we say that x and y are indistinguishable in A, in symbols xRy. The relation R partitions U into a family of disjoint subsets. Let U/R denote the quotient set consisting of equivalence classes of R. Elements of U/R are called elementary or atomic sets. The empty set and the union of one or more elementary sets are called a composed, or definable sets [13]. The family of all composed sets is denoted by Com(A). Given an arbitrary set $X \subseteq U$, in general it may not be possible to describe X precisely in A. One may characterize X by a pair lower and upper approximations.

Definition 1 *Let R be an equivalence relation on a universe U. For any set $X \subseteq U$, the lower approximation $\underline{apr}(X)$ and the upper approximation $\overline{apr}(X)$ are defined by as follows:*

$$\underline{apr}(X) = \{x \in U \mid [x]_R \subseteq X\} \quad (1)$$
$$\overline{apr}(X) = \{x \in U \mid [x]_R \cap X \neq 0\} \quad (2)$$

where

$$[x]_R = \{y \mid xRy\} \quad (3)$$

is the equivalence class containing x.

The lower approximation $\underline{apr}(X)$ is the union of elementary sets which are subsets of X, and the upper approximation $\overline{apr}(X)$ is the union of elementary sets which have a non-empty intersection with X. The set $bnd(X) = \overline{apr}(X) - \underline{apr}(X)$ is called boundary of X in A. If $bnd(X)$ is empty, then subset X is exactly definable. Note that rough set is a set of lower and upper approximation.

An accuracy measure of set X in the approximation space $A = (U, R)$ is defined as

$$\alpha(X) = \frac{|\underline{apr}(X)|}{|\overline{apr}(X)|} \quad (4)$$

where $|\cdot|$ denotes the cardinality of a set. Clearly, they are true that $0 \leq \alpha(X) \leq 1$, and $\alpha(X) = 1$ if X is definable in A; $\alpha(X) < 1$ if X is undefinable in A.

2.2 Rule extraction

A natural way to extract rules, or represent experts' knowledge, is to construct a set of conditional productions, each of them having the form

IF { set of conditions} **THEN** { set of decisions}

Such a form can be easily induced taking the advantage of rough set. In an approximation space $A = (U, R)$, regarding a subset X of U, the whole universe U is partitioned into three regions:

- Positive region $pos(X) = \underline{apr}(X)$;
- Negative region $neg(X) = U - \overline{apr}(X)$;
- Boundary region $bnd(X) = \overline{apr}(X) - \underline{apr}(X)$

which lead the following decision rules:

- Describing $pos(X) \longrightarrow$ positive decision rules;
- Describing $neg(X) \longrightarrow$ negative decision rules;
- Describing $bnd(X) \longrightarrow$ possible decision rules.

Also, the positive decision rules, possible decision rules are referred to as certain rules, possible rules, respectively. A simple illustration example is shown in Section 3 (Example 1).

2.3 Information system

In an intelligence system, the database regarding the experts' know-how is generally given in the form of the information system defined by Pawlak [7].

Definition 2 *An information systems S is an ordered quadruple*

$$S = (U, Q, V, \rho) \quad (5)$$

where U is the universe which is a non-empty finite set of objects x; Q is a finite set of attributes q; $V = U_{q \in Q} V_q$, and V_q is the domain of attribute q; ρ is a mapping function such that $\rho(x, q) \in V_q$ for every $q \in Q$ and $x \in U$. Q is composed of two parts [10]: a set of condition attributes (C) and a decision attribute (D), i.e., $Q = C \cup D$.

ρ also is called a decision function. If we introduce function $\rho_x : Q \to V$ such that $\rho_x(q) = \rho(x, q)$ for every $q \in Q$ and $x \in U$, ρ_x is called decision rule in S, and x is called a label of the decision rule ρ_x. If ρ_x is a decision rule then the restriction of ρ_x to C, in symbols ρ_x/C, and the restriction of ρ_x to D, in symbols ρ_x/D, are called conditions and decisions or actions of ρ_x, respectively.

As mentioned before, here in this paper we use the rough set to extract rules in a database, and the database is given in the form of the information system, so we need to make the connection between information system and rough set. Let $S = (U, Q, V, \rho)$ be a information system, and let $q \in Q$, $x, y \in U$. If $\rho_x(q) = \rho_y(q)$, then we say x, y are indistinguishable, in symbols $x\tilde{q}y$. Certainly, $\tilde{q}$ is an equivalence relation. Also, objects $x, y \in U$ are indistinguishable with respect to $P \subset Q$ in S, in symbols $x\tilde{P}y$, if $x\tilde{p}y$ for every $p \in P$. In particular, if $P = Q$, x, y are indistinguishable in S, in symbols $x\tilde{S}y$ instead of $x\tilde{Q}y$. Therefore each information system $S = (U, Q, V, \rho)$ defines uniquely an approximation space $A = (U, \tilde{S})$, where $\tilde{S}$ is an equivalence relation generated by the information system S. Namely, if a database is given in the form of an information system S, we can obtain an approximation space $A = (U, \tilde{S})$, further, an arbitrary subset in U is able to be approximated by the rough set.

3 Approach with GA-based Fuzzy Modeling

3.1 Problem description

Before we describe the problem, first, we show an example.

[Exmaple 1] Suppose that there is an information system $S = (U, Q, V, \rho)$, which is a database about the diagnosis of influenza (Tab.1). In the information system, $U = \{p_1, p_2, \cdots, p_6\}$ in which each object (element) expresses a patient; $Q = C \cup D = \{temp, sneeze, headache, influ\}$, $V_{temp} = \{0, 1, 2\}$ in which 0 expresses "normal", 1 expresses "high" and 2 expresses "very high"; $V_{sneeze} = V_{headache} = V_{influ} = \{0, 1\}$ in which 0 expresses "no" and 1 expresses "yes". Also, the mapping function ρ is given in the table.

Clearly, S yields the following elementary sets with respect to attributes "temp", "sneeze" and "headache":

$$E_1 = \{p_1, p_5\}, \quad E_2 = \{p_2\}, \quad E_3 = \{p_3\},$$

$$E_4 = \{p_4\}, \quad E_5 = \{p_6\}$$

i.e., $U/\tilde{S} = \{E_1, E_2, E_3, E_4, E_5\}$.

Now, let us consider to approximate a subset

Table 1: Influenza data

U	Q			
	C			D
	temp	*sneeze*	*headache*	*influ*
p_1	2	0	0	1
p_2	1	1	0	1
p_3	1	0	1	0
p_4	1	1	1	1
p_5	2	0	0	0
p_6	0	1	1	0

$$X = \{p_1, p_2, p_4\}$$

which is a set of patients who are catching a cold. Based on the concepts defined in Section 2, we have,

$$\begin{aligned} \underline{apr}(X) &= \{p_2, p_4\} \\ \overline{apr}(X) &= \{p_1, p_5, p_2, p_4\} \\ pos(X) &= \{p_2, p_4\} \\ neg(X) &= \{p_3, p_6\} \\ bnd(X) &= \{p_1, p_5\} \end{aligned}$$

therefore, $pos(X)$ follows the certain rules below:

(1) **IF** *temp*=1 $\wedge$ *sneeze*=1 $\wedge$ *handache*=0 **THEN** *influ*=1;

(2) **IF** *temp*=1 $\wedge$ *sneeze*=1 $\wedge$ *handache*=1 **THEN** *influ*=1;

where $\wedge$ denotes "and". And, $bnd(X)$ follows the possible rules below:

(3) **IF** *temp*=2 $\wedge$ *sneeze*=0 $\wedge$ *handache*=0 **THEN** *influ*=1;

(4) **IF** *temp*=2 $\wedge$ *sneeze*=0 $\wedge$ *handache*=0 **THEN** *influ*=0;

we can see that in rules (3) and (4), though they have the same condition in **IF** part, the decisions are different in **THEN** part. It means in such a case (condition), you are *probably* catching a cold. The negative decision rules are obtained by describing $neg(X)$ as follows:

(5) **IF** *temp*=1 $\wedge$ *sneeze*=0 $\wedge$ *handache*=1

THEN $influ$=0;

(6) **IF** $temp$=0 ∧ $sneeze$=1 ∧ $handache$=1 **THEN** $influ$=0;

Form (4), the approximation accuracy $\alpha(X) = 2/4 = 0.5$. □

In the above example, regarding the attribute values of $temp$, we used that $V_{temp} = \{0,1,2\} = \{"normal","high","veryhigh"\}$. Actually, the temperatures of patients are continuous. For example, temperatures of six patients (temp(p_1)∼ temp(p_6)) are given in Tab.2. In order to convert the continuous attribute values into a discrete (or digital) attribute values like $(0,1,2)$, one of the most common method is to give some "*appropriate*" intervals, each of which represents one discrete value like 1 or 2. Here, one case is shown in Fig.1. Obviously, the discrete attribute values of $temp$ in Tab.1 matches the conversion shown in Fig.1. However, if we change such a conversion like Fig.1 to Fig.2, then Tab.1 will become Tab.3 in which the (discrete) attribute values of $temp$ are different form Tab.1. In this case, regarding the same subset $X = (p_1, p_2, p_4)$,

Table 2: Real temperatures (^oC)

patient	p_1	p_2	p_3	p_4	p_5	p_6
temp	40.1	37.5	37.8	38.0	38.3	36.5

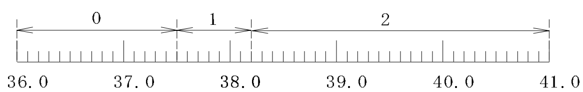

Figure 1: One case of conversion

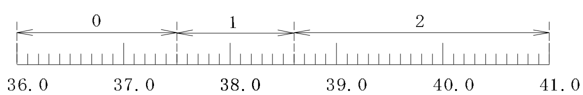

Figure 2: Another case of conversion

$$\underline{apr}(X) = \{p_1, p_2, p_4\}$$
$$\overline{apr}(X) = \{p_1, p_2, p_4\}$$
$$\alpha(X) = 1$$

Clearly, the approximation accuracy is improved. Therefore, even with a same original database in which continuous attribute values are contained, the different conversion has a different approximation accuracy, which directly influence the rules extracted from the database. Consequently, when we consider such a conversion for the originally continuous attribute values, there are two problems we have to answer:

(1) How to divide the continuous values into some intervals, each of which corresponds to a discrete number?

(2) How many intervals should be taken?

Table 3: Influenza data with different conversion

U	Q			
	C			D
	$temp$	$sneeze$	$headache$	$influ$
p_1	2	0	0	1
p_2	1	1	0	1
p_3	1	0	1	0
p_4	1	1	1	1
p_5	1	0	0	0
p_6	0	1	1	0

3.2 Proposal of solution

The problem (1) comes to how to give a proper boundary between two neighboring intervals. Here, we recall the fuzzy set theory, which deals with the ill-definition of the boundary through so-called fuzzy membership functions. First, we make a connection between the fuzzy set and usual conversion like Fig.2. For example, as shown in Fig.3, fuzzy sets $\tilde{0}, \tilde{1}, \tilde{2}$, and $\tilde{3}$ correspond to intervals $0, 1, 2$, and 3, respectively. Consider a real attribute value x in Fig.3, because their membership values relation of $\mu_{\tilde{1}}(x) > \mu_{\tilde{0}}(x)$, so x is reasonably divided into

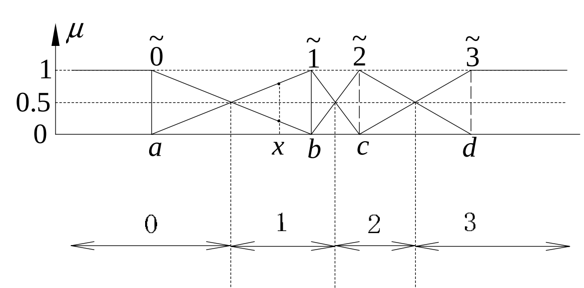

Figure 3: Correspondence between fuzzy sets and usual division

interval 1.
In this paper, we use triangular fuzzy memberships, and the location of the corresponding fuzzy membership functions are chosen so that they always overlap at the membership value $\mu = 0.5$, therefore the fuzzy membership functions, say number is n, can be determined completely, if n parameters like a, b, c, and d are determined. Such a process is called fuzzy modeling. To do so, again, let us recall the genetic algorithm (GA), which has been widely used in various problems as a robust search method, especially in optimum seeking. Also, as mentioned below, the problem (2) in the previous sub-section can be resolved in passing if we use GA. In the upcoming sub-section, we will describe the GA-based fuzzy modeling in detail.

3.3 GA-based fuzzy modeling

As an important branch of evolutionary computation (EC), genetic algorithm (GA) is characterized by its current effectiveness, strong robustness, and simple implementation. It also has the advantage of not being restrained by certain restrictive factors of search space. GA simulates the evolutionary process of a set of "genomes" over time. Genome is a biological term that refers to a set of "genes" and gene is the basic building block of any living entity. For our use now, "genome" represents two-figure hexadecimal such as "A8", which finally can be transferred to the values of parameters like a, b, c in Fig.3, and gene represents the binary digit in the binary coded hexadecimal code (BCHC) such as "10101000", which is the BCHC of "A8". A GA starts with a set of genomes, which is referred to as a generation, created randomly and then the evolutionary process of the "survival of the fittest" genomes takes place. The unfitted genomes are removed and the remaining genomes reproduce a new set of genomes. Reproduction of the genomes is accomplished by applying the simulation of the two well-known genetic processes: mutation and crossover. This process is repeated and in each repetition a fitter generation is created. To fit the use in this paper, our GA is composed of the following steps:

Step 1. Find the biggest value x_{max} and the smallest value x_{min} in the continuous attribute values x to be considered, and evenly divide $[x_{min}, x_{max}]$ into seven interval so that 8 points correspond to a BCHC.
In this way, we have built a connection between a genome and a fuzzy division. Actually, the 8 points are candidates of division points. Binary digit 1 in BCHC means that it is a division point, and 0 means that it is not a division point. For example, a hexadecimal "4C" leads a fuzzy division shown in Fig.4.

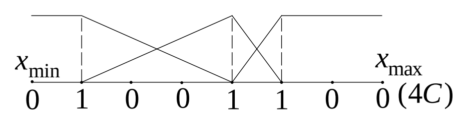

Figure 4: Fuzzy division of a BCHC

Step 2. Create a set of N random BCHC (first generation of genomes).

Step 3. Calculate the fitness of each BCHC.
Each BCHC leads an information system like Tab.1 or Tab.3, therefore for a subset X to be approximate, there are different approximation accuracy $\alpha(X)$ in (4). Here we use $\alpha(X)$ as the fitness. Naturally, a better accuracy presents a better fitness.
In Fig.4, the interval number is 3, which equals the sum of binary digit 1 in BCHC(4C). We may suppose that the maximum number $Itvl_{max}$ of intervals is given. So when we calculate the fitness of each BCHC, we will give the worst fitness, say 0, if the sum of binary digit 1 in a BCHC exceeds $Itvl_{max}$.
GA goes to the end if the desired fitness is obtained. At the same time, the problem (2) described in previous sub-section is resolved, namely, the sum of binary digit 1 in a BCHC with the best fitness is interval number we should take.

Step 4. Sort the BCHCs based on their fitness in descending order.

Step 5. Keep M ($M < N$) fitter BCHCs and remove the rest of the BCHCs.

Step 6. Create the next generation by making $N - M$ BCHCs out of M BCHCs using crossover and mutation operations.

Step 7. Go to step 3.

4 Conclusion

The main purpose in this paper is to effectively extract rules from a database which is given in the form of the information system based on the rough set theory. To do so, the key point is how to approximate a subset with the best accuracy. In this paper, we proposed an approach for rule extraction using the GA-based fuzzy modeling. Hereafter, we will apply our approach to a home-helper-oriented medical support system.

References:

[1] L. A. Zadeh, "Fuzzy set," *Information and Control*, vol. 8, pp.338-353, 1965

[2] P. J. King, E. H. Mamdami, "The application of fuzzy control systems," In M. M. Gupta, G. Saridis and B. Gaines, Eds. *Fuzzy Automata and Decision Processes*, pp.321-330, Amesterdam: North Holland, 1975

[3] T. Takagi, M. Sugeno, "Fuzzy Identification of systems and its application to modeling and control," *IEEE Transaction on Systems, Man, and Cybernetics*, vol. SMC-15, no. 1, pp.116-132, 1985

[4] L.-X. Wang, "Stable adaptive fuzzy control of nonlinear system," *IEEE Trans. on Fuzzy Systems*, vol. 1, pp. 146-155, 1993.

[5] H. Han!$C.-Y. Su, and Y.Stepanenko,"Adaptive control of a class of nonlinear systems with nonlinearly parameterized fuzzy approximators," *IEEE Transactions on Fuzzy Systems,* vol. 9, no. 2, pp. 315-323, 2001

[6] Z. Pawlak, "Rough sets," *Int. J. of Information and Computer Sciences*, vol. 11, no. 5, pp. 341-356, 1982

[7] Z. Pawlak, "Rough classification," *Int. J. of Man-Machine Studies*, vol. 20, pp. 469-483, 1984

[8] S.K.M. Wang, and W. Ziarko, "Comparison of the probability approximate classification and the fuzzy set model," *Fuzzy Sets and Systems*, 21, pp.357-362, 1987

[9] Z. Pawlak, S.K.M. Wang, and W. Ziarko, "Rough sets: probability versus deterministic approach," *Int. J. Man-Machine Stud.*, 29, pp.81-95, 1988

[10] A. Mrozek, "Rough sets and dependency analysis among attributes in computer implementations of expert's inference models," *Int. J. of Man-Machine Studies*, vol. 30 , pp.457-473, 1989

[11] H. Tanaka, H. Ishibuchi, and N. Matsuda, "Reduction of information systems based on rough sets and its application to fuzzy expert systems," *Transactions of the Society of Instrument and Control Engineerings*, vol. 27, no. 3, pp.357-364, 1991 (in Japanese)

[12] R. R. Hashemi, F. R. Jelovsek, "Inductive learning from examples: a rough sets approach," *Proc. of The 1992 ACM Conference on Applied Computing*, pp.346-349, 1991

[13] Y. Yao, "Two views of the theory of rough sets in finite universes," *Int. J. of Approximate Reasoning*, 15, pp. 291-317, 1996

[14] T. Y. Lin, and N. Cercone (Eds.), *Rough Sets and Data Mining (Analysis of Imprecise Data)*, Kluwer Academic Publishers, 1997

[15] N. Zhong, A. Skowron, and S. Ohsuga (Eds.), *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (LNAI 1711)*, Springer, 1999

[16] Z. Pawlak, "Rough sets and decision algorithms," *Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing*, pp.1-16, 2000

[17] Y. Yao, "A comparative study of fuzzy sets and rough sets," *J. of Information Sciences*, 109, pp. 227-242, 1998