Rough Sets and Kowledge Acquisition

Hideo Tanaka
Osaka Prefecture University
tanaka@ie.osakafu-u.ac.jp

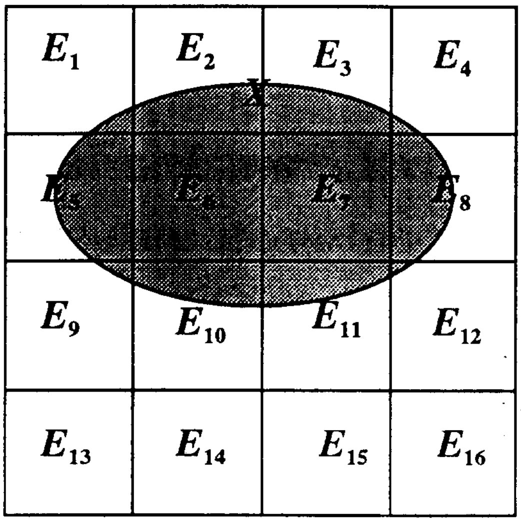
1. Introduction

Rough sets proposed by Z. Pawlak in 1982 can be described as approximate inclusion of sets. Rough sets concept can be applied to automatic classification, pattern recognition, learning algorithms, etc. Especially, approximate classification has been developed and applied for medical diagnosis [1].

[1] Z. Pawlak: Rough classification, Int. J. Man-Machine Studies 20, 469-483, 1984

2. Rough Sets Concepts

An approximation space A is the ordered pair $A = (U, R)$ where U is the universe and R is an indiscernibility (equivalence) relation. If $(x_1, x_2) \in R$, x_1 and x_2 are indiscernible in A. Equivalence classes of the relation R are called elementary sets. Any finite union of elementary sets in A is called a definable sets in A.



$U/R = \{E_1, \ldots, E_{16}\}$

Fig.1 An approximation space A.

$\{E_1, \ldots, E_{16}\}$ are elementary sets.

The upper approximation of X:

$$A^*(X) = \{E_i | E_i \cap X \neq \emptyset\} = \{E_1, \ldots, E_{12}\} = E_1 \cup \cdots \cup E_{12}$$

(the least definable set containing X)

The lower approximation of X:

$$A_*(X) = \{E_i | E_i \subset X\} = \{E_6, E_7\} = E_6 \cup E_7$$

(the greatest definable set contained X)

3. Information Tables and Equivalenve Classes

Table 1. Information systems (Z. Pawlak)

Patient	Headache	Musclepain	Temperature	Flu	
x_1	no	yes	high	yes	$q_1(x_1)= 0$
x_2	yes	no	high	yes	$q_1(x_2)=1$
x_3	yes	yes	very high	yes	
x_4	no	yes	very high	yes	⋮
x_5	no	yes	normal	no	
x_6	yes	no	high	no	$q_1(x_6)= 1$

↓

U	q_1	q_2	q_3	d_1
x_1	0	1	1	X_1
x_2	1	0	1	X_1
x_3	1	1	2	X_1
x_4	0	1	2	X_1
x_5	0	1	0	X_2
x_6	1	0	1	X_2

normal: 0
high: 1
very high: 2

$U=\{x_1,\ldots,x_6\}$ $Q=\{q_1,q_2,q_3\}$ $D=\{d_1\}$
(the universe) (the set of attributes) (the set of decisions)

q_1: $E_1(q_1)=\{x_1,x_4,x_5\}$, $E_2(q_1)=\{x_2,x_3,x_6\}$
q_3: $E_1(q_3)=\{x_1,x_2,x_6\}$, $E_2(q_3)=\{x_3,x_4\}$, $E_3(q_3)=\{x_5\}$

Q: $E_1(Q)=\{x_1\}$, $E_2(Q)=\{x_2,x_6\}$, $E_3(Q)=\{x_3\}$
$E_4(Q)=\{x_4\}$, $E_5(Q)=\{x_5\}$ $\cdots$ U/Q

$A_*(X_1)=\{E_1(Q),E_3(Q),E_4(Q)\}=\{x_1,x_3,x_4\}$
$A_*(X_2)=\{E_5(Q)\}=\{x_5\}$

4. Reduction and Decision Rules

Accuracy measure (the guality of the classification)

$$\beta(Q) = \frac{card(A_*(X_1) \cup A_*(X_2))}{card(U)}$$

where $card(U)$ is the number of elements in U.

$$\beta(Q) = 4/6 = 2/3$$

$Q'=\{q_1,q_3\}$: $E_1(Q')=\{x_1\}$, $E_2(Q')=\{x_2,x_6\}$, $E_3(Q')=\{x_3\}$

$E_4(Q')=\{x_4\}$, $E_5(Q')=\{x_5\}$ $\cdots$ $U/Q'=U/Q$

$$\beta(Q') = 2/3 = \beta(Q),\ but\ Q' \subset Q$$

- q_2 is called superfluous in Q.
- Q is said to be dependent iff there exists $Q' \subset Q$ such that $\beta(Q') = \beta(Q)$.

q_3: $A_*(X_1)=E_2(q_3)=\{x_3,x_4\}$, $A_*(X_2)=E_3(q_3)=\{x_5\}$

$$\beta(q_3) = 3/6 = 1/2 \leq \beta(\{q_1,q_3\})$$

q_1: $A_*(X_1)=\varnothing$, $A_*(X_2)=\varnothing$

$$\beta(q_1) = 0/6 = 0 \leq \beta(\{q_1,q_3\})$$

- $\{q_1,q_2\}$ is said to be independent in Q.
- Q is said to be independent iff for every $Q' \subset Q$, $\beta(Q') < \beta(Q)$.
- If $\beta(P') < \beta(P) = \beta(Q)$ for $P' \subset P \subset Q$, P is said to be reduct of Q.

In the example, $Q'=\{q_1,q_3\}$ and $Q''=\{q_1,q_2\}$ are reducts of Q.

U Patient	q_1 Headache	q_3 Temperature	d_1 Flu
x_1	no	higu	yes
x_2	yes	high	yes
x_3	yes	very high	yes
x_4	no	very high	yes
x_5	no	normal	no
x_6	yes	high	no

Reduction of Attributes

U/q_3: $E_1(q_3)=\{x_1,x_2,x_6\}$, $E_2(q_3)=\{x_3,x_4\}$, $E_3(q_3)=\{x_5\}$

U/d_1: $E_1(d_1)=\{x_1,x_2,x_3,x_4\}$, $E_2(d_1)=\{x_5,x_6\}$

$$\rightarrow E_2(q_3) \subset E_1(d_1),\ E_3(q_3) \subset E_2(d_1)$$

U Patient	q_1 Headache	q_3 Temperature	d_1 Flu
x_1	no	higu	yes
x_2	yes	high	yes
x_3	—	very high	yes
x_4	—	very high	yes
x_5	—	normal	no
x_6	yes	high	no

(—: do not care)

Reduction of Values of Attributes

if (Temprature, very high), then (Flu, yes)
if (Temprature, normal), then (Flu, no)
if (Headache, no) and (Temprature, high), then (Flu, yes)
if (Headache, yes) and (Temprature, high), then (Flu, yes)
if (Headache, no) and (Temprature, high), then (Flu, no)

Let $F=\{X_1,...,X_n\}$ be a classification of U, i.e. $X_i \cap X_j \neq \varnothing$ for every $i \neq j$ and $\cup_{i=1}^{n} X_i = U$. Then F is called a partition of U and X_i is called a class. An upper approximation and lower approximation of$F$ can be written as:

$$A^*(F)=\{A^*(X_1),...,A^*(X_n)\}$$

$$A_*(F)=\{A_*(X_1),...,A_*(X_n)\}$$

An accuracy measure of F in U is defined as:

$$\beta_A(F) = card(\cup_{i=1}^{n} A_*(X_i)) / card(U)$$

Definition 1:

i) Let P be a subset of Q. A subset P is said to be independent if a nd only if

$$\beta_A(P') < \beta_A(P) \quad for\ all \quad P' \subset P.$$

Also, P is said to be dependent if and only if there is $P' \subset P$ such that $\beta_A(P') = \beta_A(P)$.

ii) Let $P' \subset P$ and $P'' = P - P'$. A subset P' is said to be superfluous in P if and only if $\beta_A(P'') = \beta_A(P)$.

Algorithm

Step 0: Set $P = Q$.

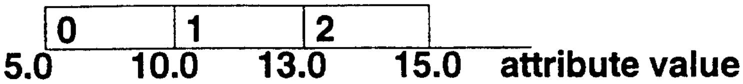
Step 1: Find a superfluous attribute, say p_i, in P. If there is not such a p_i go Step 4.

Step 2: Set $P = P - \{p_i\}$.

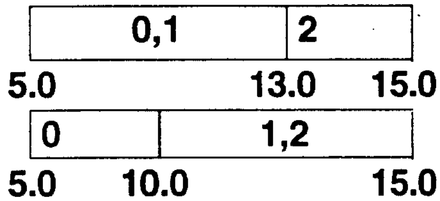
Step 3: If all $\{p_i\} \subset P$ is not superfluous in P, go to Step 4. Otherwise, go to Step 1.

Step 4: End. P is the reduct of the given attributes.

5. Reduction of Divisions of Attributes [2]



(Medical inspection)



A binary representation of {0,1,2}

	z_1	z_2
0	0	0
1	0	1
2	1	1

Table 2 . An example of information system

U	q_1	q_2	q_3	q_4
x_1	0	2	1	0
x_2	0	1	2	1
x_3	1	0	1	0
x_4	1	0	1	0
x_5	1	2	1	0
x_6	1	2	0	0

Table 3. A binary representation of Table 2

F	U	q_1	q_2		q_3		q_4
		z_1	z_2	z_3	z_4	z_5	z_6
X_1	x_1	0	1	1	0	1	0
	x_2	0	0	1	1	1	1
	x_3	1	0	0	0	1	0
	x_4	1	0	0	0	1	0
	x_5	1	1	1	0	1	0
	x_6	1	1	1	0	0	0

Table 4. The reduction result

	q_1	q_2	q_3
x_1	0	2	1,2
x_2	0	0,1	1,2
x_3	1	0,1	1,2
x_4	1	0,1	1,2
x_5	1	0,1	1,2
x_6	1	0,1	0

[2] H. Tanaka, K. Koyama and Y. Maeda: A Method for Reducing Information Systems with Binary Data by Rough Sets, Fifth IEEE Int. Conference on Fuzzy Systems, September 1996

6. Application [3]

In order to show that this reduction method in the section 2 is useful, we applied our method to **medical data analysis for hepatic diseases**. These data consist of 5 classes.

Healthy person : X_1
Hepatoma : X_2
Acute hepatics : X_3
Chronic hepatics: X_4
Liver cirrhosis : X_5

The number of medical inspection is 20. The integers of attribute values are given by medical experts as shown in Table. A default value is represented as 0 in the data. The given data from **Kawasaki Medical College** are as follows:

$U = \{x_i\}, i = 1,\ldots,468$; 468 samples

$Q = \{q_j\}, j = 1,\ldots,20$; 20 medical inspections

$F = \{X_k\}, k = 1,\ldots,5$; 5 classes

[3] H.Tanaka, H. Ishibuchi and N. Matsuda: Fuzzy Expert System Based on Rough Sets and Its Application to Medical Diagnosis, Int. J. of General Systems, Vol.21, pp.83-97, 1992

Examples of divisions of medical test data

	Medical Inspection	Integers of attribute values					
		1	2	3	4	5	6
q_1	SP	~5.5	5.6-6.5	6.6-7.5	7.6-8.5	8.6~	***
q_2	II	~4	5-6	7-9	10~	***	***
q_9	ChE	~100	101-150	151-200	201-250	251-500	501~
q_{10}	GPR	~25	26-100	101-200	201-500	501-1000	1001~
q_{16}	Lympho	~20.0	20.1-40.0	40.1-60.0	60.1~	***	***
q_{19}	A1-%	~2.5	2.6-3.7	3.8-5.0	5.1~	***	***
q_{20}	AFP	~20	21-100	101-200	201-1000	1001~	***

Binary representation of 7 attributes

Division	Attribute											
	q_2, q_{16}, q_{19}			q_1, q_{20}				q_9, q_{10}				
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	1	0	0	0	0	1
3	0	1	1	0	0	1	1	0	0	0	1	1
4	1	1	1	0	1	1	1	0	0	1	1	1
5				1	1	1	1	0	1	1	1	1
6								1	1	1	1	1

First Step							Second Step					
1	2	3	4	5		Attribute 1	1	2	3	4	5	
1	2	3	4			Attribute 2	1	2	3	4		
1	2	3	4	5	6	Attribute 9	1	2	3	4	5	6
1	2	3	4	5	6	Attribute 10	1	2	3	4	5	6
1	2	3	4			Attribute 16	1	2	3	4		
1	2	3	4			Attribute 19	1	2	3	4		
1	2	3	4	5		Attribute 20	1	2	3	4	5	

Total number of divisions 34 Total number of divisions 25

Fig. 2. The reduction result by our method

7. Fuzzy if-then rules

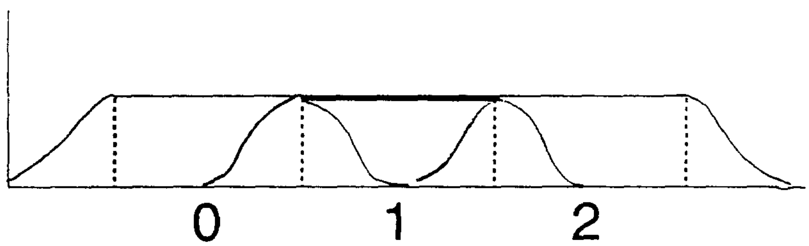
From Table 1, we have the following consistent if-then rules:

If x is $q_2=0$ and $q_3=1$, then x is $d_1=1$ (Flu).

If x is $q_3=2$, then x is $d_1=1$.

If x is $q_3=0$, then x is $d_1=0$.

q_2 \ q_3	0	1	2
1	$d_1=0$		$d_1=1$
0	$d_1=0$	$d_1=1$	$d_1=1$



Fuzzy intervals by exponential functions

8. Similar concepts to Rough Sets

(1) Retrieval in incomplete information

Let us get started with an example of information retrieval in the database of incomplete information shown in Table 5 where it is assumed that the age of a person is exactly unknown, but given by the interval of possible ages. It should be noted that intervals are the simplest one among fuzzy numbers. In addition, we let the age of the person being searched for the interval Q=[20,25]. Then, we have the incomplete information and also the incomplete query. There are the limiting interval interpretation of the query Q from the two viewpoints of possibility and necessity as follows [4]:

(i) The upper value of querr, $A^*(Q)$, which is the set of objects from which the available information could possibly satisfy the query Q **[possibility]**.

(Ii) The lower value of querr, $A_*(Q)$, which is the set of objects from which the available information can not fail to satisfy the query Q **[necessity]**.

Table 5. An example of database of incomplete information

Name	Age
a	X_a=[23,26]
b	X_b=[20,22]
c	X_c=[30,36]
d	X_d=[20,23]
e	X_e=[27,31]

More specifically, (i) and (ii) can be translated into the following mathematical expressions, respectively:

$$A^*(Q) = \{i|\ X_i \cap Q \neq \emptyset\} = \{a,b,d\} \quad (possibility)$$

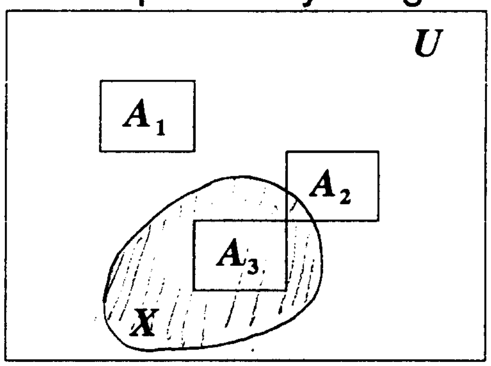
$$A_*(Q) = \{i|\ X_i \subseteq Q\} = \{b,d\} \quad (necessity)$$

[4] W. Lipski: On semantic issues connected to incomplete information in databases, ACM Trans. on Database Systems, 4, 262-296, 1979

(2) A Mathematical Theory of Evidence [5]

[5] G. Shafer: Princeton Univ. Press, 1976
[6] A. P. Dempster: Upper and lower probabilities induced by a multivalued mapping, Annals. of Math. Stat. 38, 2, 325-339, 1967

A basic probability assignment (a structure of evidence)



$m(A_1) = \mathbf{0.3}$

$m(A_2) = \mathbf{0.5}$

$m(A_3) = \mathbf{0.1}$

$m(U) = \mathbf{0.1}$

Total 1.0

The upper probability

$$P^*(X) = \sum_{A_i \cap X \neq \emptyset} m(A_i) = 0.5 + 0.1 + 0.1 = 0.7 \ (\textit{possibility})$$

The lower probability

$$P_*(X) = \sum_{A_i \subseteq X} m(A_i) = 0.1 \ (\textit{necessity})$$

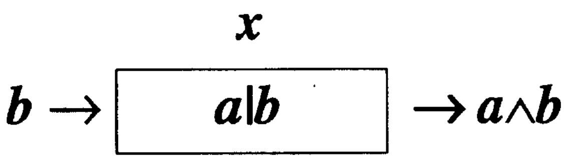
Example: The Ming Vase: I contemplete a vase that has been represented as a product of the Ming dynasty. Is it genuine or counterfeit?

θ_1: The vase is genuine, θ_2: it is counterfeit

$$m(\{\theta_1\}) = 0.4, \ m(\{\theta_1, \theta_2\}) = 0.6$$

$$P^*(\theta_1) = 0.4 + 0.6 = 1.0, \ P_*(\theta_1) = 0.4$$

(3) Conditional Logic [7]



$$a|b = \{x \mid x \wedge b = a \wedge b\}$$

$$= [a \wedge b, \neg b \vee a]$$

$$a \wedge b \subseteq \neg b \vee a$$

$A_* \subseteq A^*$ (Rough Sets)

$a|b$: if x is b, then y is a. (conditional logic)

$b \to a$: if x is b, then y is a. (Implication)

$$b \to a = \sup_x \{x \mid x \wedge b \leq a\}$$

$$= \neg b \vee a$$

[7] Conditional Logic in Expert Systems: Edited by I. R. Goodman et al., North-Holland, 1991

(4) Interval Regression Analysis [8]

Given data: $(Y_j, \boldsymbol{X}_j)$, j=1,....,m, $\boldsymbol{X}_j$=(X_{j1},...,X_{jn})

$$Y_j^* = A_1^* x_{j1} + \cdots + A_n^* x_{jn} \quad (possibility\ model)$$

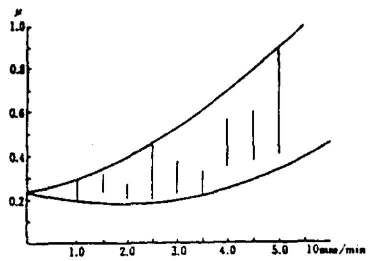
$$Y_{*j} = A_{*1} x_{j1} + \cdots + A_{*n} x_{jn} \quad (necessity\ model)$$

$$A = (\boldsymbol{a}_c, \boldsymbol{a}_w)$$

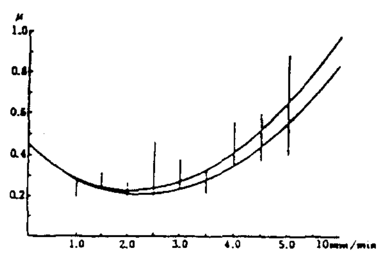
i) $Y_j \subset Y_j^*$, $J^* = \Sigma \boldsymbol{a}_w^{*t} |\boldsymbol{x}_j| \rightarrow \min$: *Possibility model*

ii) $Y_{*j} \subset Y_j$, $J_* = \Sigma \boldsymbol{a}_{*w}^{t} |\boldsymbol{x}_j| \rightarrow \max$: *Necessity model*

[8] H.tanaka, I. Hayashi and J. Watada: Neccesity model regression analysis for fuzzy data, Europian J. of Oper. Res., 40, 389-396, 1989



Possibility estimation model



Necessity estimation model

Table 2. Feed speed and surface roughness

		Roughness(μ)y_i		
No.	Feed speed x (10 mm/min)	min. value	max. value	Roughness(μ)Y_i (y_i, e_i)
1	1.0	0.19	0.29	(0.240,0.050)
2	1.5	0.24	0.32	(0.280,0.035)
3	2.0	0.20	0.27	(0.235,0.035)
4	2.5	0.20	0.46	(0.330,0.130)
5	3.0	0.22	0.38	(0.300,0.080)
6	3.5	0.22	0.33	(0.275,0.055)
7	4.0	0.35	0.56	(0.455,0.105)
8	4.5	0.37	0.60	(0.485,0.115)
9	5.0	0.41	0.89	(0.650,0.240)

9. Concluding Remarks

1) Rough Sets concept is similar to Fuzzy Set Concept.

2) Our Partial ignorance can be expressed by intervals.

3) Knowledge acquisition will be obtained by Rough Sets.

4) There are many research topics related to Rough Sets Concepts such as Interval Regression.