ORIGINAL ARTICLES

KNOWLEDGE ACQUISITION FROM THE DATA CONSISTING OF CATEGORIES ADDED WITH DEGREES OF CONFORMITY

Proposal of Extended Reduct Calculation in Rough Set Theory

Norihiko MORI, Rei TAKANASHI

Tokyo Institute of Polytechnics, 2-9-5 Honmachi, Nakano-ku, Tokyo-to, 164-8678, Japan

Abstract: Reduct of the rough set theory is recently used as a mean of knowledge acquisition concerned to a cause-effect relationships consisting of attributes of an object as a cause and of classification as an effect. The data of the rough set theory is limited to categorical data. However, *kansei* data are often ambiguous. Accordingly, a certain observation value might not conform to a certain category but might only conform to within some grade. In order to make the knowledge acquisition be more agreeable to the actual condition, this paper proposes an algorithm of reduct calculation for the categories added with degrees of conformity which is appropriate to the expression of attributes related to *kansei*.

Keywords: Reduct, Discernibility matrix, Rough set theory, Knowledge acquisition, Category with grade

1. BACKGROUND AND PURPOSE

Data mining and knowledge acquisition are approaches for finding certain rules from the inside of complicated data concerning cause-effect relationships. In recent years, the rough set theory has come to stand out as one of the successful means in these fields. Therefore, the application of the rough set theory is increasing in some other fields including medicing [1-3].

The relationship between the attributes of an object and the classification of the object form a kind of cause-effect relationship. However, if attributes come to consist of *kansei* related words, or classification comes to be done with some *kansei*, the data of the cause-effect relationship comes to be ambiguous as an inevitable consequence.

Some case studies related to knowledge acquisition have the purpose of support for design, and these studies execute reduct calculations, which make up one of the main tools in the rough set theory, as the means of knowledge acquisition. In these studies, attributes consist of form elements of products, whose categories are judged cognitively, and classification consists of image evaluation [4-6].

When the rough set theory is applied, the data is limited to categorical data. However, data related to *kansei*, i.e., cognition, image evaluation, etc., is known to often be ambiguous. Accordingly, a certain observation value might not conform to a certain category but might only conform to within some grade. Therefore, it seems to appropriate to extend the rough set theory so that it can be applied to grade added categorical data.

The purpose of this paper is to develop an algorithm for finding reduct and to show the usefulness of the reduct for knowledge acquisition when the data about the cause consists of categories with grades and the data about the effect consists of ordinary categories.

2. CONCEPT, DEFINITION, AND DATA

A grade is not a probability. A grade is an expression of the degree to which an observation value conforms to a category. It is synonymous with the membership value in the fuzzy set theory.

Expressing the grade as p , we get

$$p=[0, 1]$$

In the fuzzy set theory, grade p in category A is written as p/A , but in this paper, this case is simply written as

$$pA$$

The difference $pA \sim qA$ between grade p and grade q is defined as

$$pA \sim qA = (p \sim q)A$$

this equation shows the degree of the difference.

In order to make acquired knowledge as effective as possible, the knowledge concerning the cause is expressed with not a certain grade but an interval of a grade.

For example, the knowledge that "If an attribute has 0.8A (i.e., A of grade 0.8) then belonging to a certain class is effected with a degree of 0.6" leads to more broad and

effective knowledge that “ If an attribute has 0. 8A or more (i.e., A of grade 0. 8 or more) then belonging to a certain class is effected with a degree of 0. 6 at least ”.

Here, “0. 8A or more “ denotes an interval of a grade.

Table 1 is an example of categorical data. Table 2 gives the data of Table 1 transferred into grade added categorical data. Table 3 shows the transfers. This paper adopts Table 2.

Table 1: Categorical data

	attributes					class
	1	2	3	4	5	Y
U1	A	C	E	G	J	1
U2	B	C	E	H	J	2
U3	A	C	F	H	K	2
U4	A	D	F	G	K	2

Table 2: Grade added categorical data

	categories					class
	A	C	E	G	J	Y
U1	0. 7A	0. 8C	0. 6E	0. 9G	1. 0J	1
U2	0. 3A	0. 8C	0. 7E	0. 2G	1. 0J	2
U3	0. 6A	0. 8C	0. 2E	0. 2G	0. 0J	2
U4	0. 7A	0. 4C	0. 2E	0. 7G	0. 0J	2

Table 3: Transfer of data

Table 1 ----->Table 2	
A----->	[0. 5, 1]A
B----->	[0, 0. 4]A
C----->	[0. 5, 1]C
D----->	[0, 0. 4]C
E----->	[0. 5, 1]E
F----->	[0, 0. 4]E
G----->	[0. 5, 1]G
H----->	[0, 0. 4]G
J----->	[0. 5, 1]J
K----->	[0, 0. 4]J

In order to clarify the properties of reducts led from grade added categorical data, both reducts led from Table 1 and reducts led from Table 2 are compared.

In Table 2, A, C, E, G, and J are categories of attributes assumed as causes, U denotes an object, and Y denotes a class assumed as the result.

In the real world, U can be a car, for example, and the categories of attributes can be:

A : A personality-shaped head lamp

C : A silhouette with a strong trapezoidal degree

E : A big and conspicuous tire

G : A car low in height.

J : A showy body color

Y=1: Sporty

Y=2 : Not sporty.

Here, we discuss of the method of transferring from the observation data in the real world into grade added categorical data.

The attributes of an object assumed as the cause can include, for example, “roundness of whole shape”, and supposing that it can be divided into five categories, i. e., very round / rather round / middle / rather angular / very angular, it can be assigned five grades, i. e., 0.9 / 0.7 / 0.5 / 0.3 / 0.1 in the order of categories.

Here, we consider that the attribute of “roundness of whole shape” can be expressed with only one category with grades having an interval measure.

However, if the attribute is “vividness of color” and supposing that has two categories, i. e., vivid / quiet , it can hardly be expressed with only one category. This is because the vividness of a certain color occasionally has two categories, i. e., one is vivid which has a grade of 0.9 , and the other is quiet which has a grade 0.5. In this case, the two grades do not add up to 1.

3. DISCERNIBILITY MATRIX

Next, the discernibility matrix for grade added categorical data is formed from the data for each Y of the result.

Table 4: Discernibility matrix for categorical data

Y=1				Y=2	
U	2	3	4	U	1
1	A	E	C	2	B
	G	G	E		H
		J	J	3	F
					H
					K
				4	D
					F
					K

This matrix is equivalent to the discernibility matrix for ordinary categorical data, which is used in Shan/Ziarko's algorithm [7].

Table 4 is a discernibility matrix formed from the categorical data of Table 1 [8].

The following is an explanation of a process for forming a discernibility matrix for categorical data with grades.

We take a discernibility matrix for Y=1 as an example.

Every U classified to Y=1 is put in a row of the matrix, and every U classified to other Ys is put in a column of the matrix. All of the Us in the rows and all of the Us in the columns are compared with each other for every category in the data.

If there is a difference between U1(row) and U2(column) of column for category A , the difference value, i. e., the result obtained by subtracting the grade of category A of U2 from the grade of category A of U1, is put as an element of the matrix with a plus sign or minus sign.

We can achieve the purpose of finding an interval of reduct by means of considering that, when the value has a plus sign, the grade of category A of U1 means not only the grade shown in the data but also the grade or more too, and when the value has a minus sign, the grade of category A of U1 means not only the grade shown in the data but also the grade or less too.

When the subtraction result of a certain category, e.g. A , has a plus sign, A is denoted by A, and when it has a minus sign, A is denoted by A[~].

Table 5: discernibility matrix with grades

Y=1			
U	2	3	4
1	0. 4A	0. 1A	0. 4C
	0. 1E [~]	0. 4E	0. 4E
	0. 7G	0. 7G	0. 2G
		1. 0J	1. 0J

Y=2	
U	1
2	0. 4A [~]
	0. 1E
	0. 7G [~]
3	0. 1A [~]
	0. 4E [~]
	0. 7G [~]
	1. 0J [~]
4	0. 4C [~]
	0. 4E [~]
	0. 2G [~]
	1. 0J [~]

The discernibility matrices for the other Ys are formed in a like manner.

Table 5 is a discernibility matrix formed from Table 2. Its element shows a degree of difference between grades.

We note that an element of a discernibility matrix with grades represents a difference based on a grade added category of U in a row.

If we explain this by an example using Table 5 , 0. 4A of row 1 and column 1 represents the case that if A of U1 is 0. 7Aor more then it has a difference of 0. 4 at least for A of U2, and 0. 1E[~] of row 2 and column 1 represents the case that if E of U1 is 0. 6E or less then it has a difference of 0. 1 at least for E of U2.

Table 5 returns to Table 4 when every [0.5, 1] in Table 3 is replaced with 1 and every [0, 0.4] is replaced with 0. Therefore, a grade added discernibility matrix is an extended version of the discernibility matrix in Shan/Ziarko's algorithm.

4. REDUCT CALCULATION AND KNOWLEDGE EXPRESSION

The minimal reduct in the rough set theory is the minimal and sufficient condition for attributes such that an object belonging to a certain class can be discerned from among all of the objects belonging to all of the other classes[9].

In knowledge acquisition for the purpose of design support, there is a need for the acquired knowledge to be accurate. Accordingly, we try not to find a reduct of possibility but to find a reduct of necessity, that is, the minimal reduct.

The minimal reduct is obtained by a boolean operation on the or-associations among the rows and the and-associations among the columns in the discernibility matrix. Hereafter, the minimal reduct is simply described as the reduct.

The boolean operation is described as follows using or-associations denoted by + , and and-associations denoted by *.

$$A * A = A, \quad A + A = A, \quad A + AB = A (1 + B) = A$$

At first, the boolean operation for Y=1 in Table 4 (i. e., a discernibility matrix of ordinary categorical data) is executed and five reducts AE, AJ, GC, GE, and GJ are obtained as shown below.

$$(A + G) (E + G + J) (C + E + J) = AE + AJ + GC + GE + GJ \text{-----} (1)$$

Next, the boolean operation for a discernibility matrix of grade added categorical data is defined as follows.

Because the binary operation between degrees of difference of grades is a logical operation, a + sign means $\vee$ (taking the bigger one) and * means $\wedge$ (taking the smaller one). Therefore, the binary operation between the degrees of difference of grades in the same categories is defined as below (s and t denote the degrees of difference).

$$\left. \begin{array}{l} sA + tA = (s \vee t)A \\ sA * tA = (s \wedge t)A \\ \text{and } sAA = sA \end{array} \right\} \text{-----(2)}$$

The following is noteworthy.

$$AA \sim \neq A$$

The binary operation in the different categories is given as below.

When $s \geq t$

$$sA + tAB = sA \text{ -----(3)}$$

This is because, when $s \geq t$

$$sA + tA = (s \vee t)A = sA$$

Therefore, extending $A + AB = A$ in the Boolean operation for the ordinary categorical data

$$sA + tAB = sA + tA + tAB = sA + tA = sA$$

It is noteworthy that when $s < t$, equation (3) is not true.

4.1. Reduct calculation in the case of a single object

In Table 2, U classified to $Y=1$ is only U1.

By applying (2) and (3) to $Y=1$ in Table 5 of the discernibility matrix

$$\begin{aligned} & (0.4A + 0.1E \sim + 0.7G) * (0.1A + 0.4E + 0.7G + 1.0J) * (0.4C + \\ & 0.4E + 0.2G + 1.0J) \\ & = 0.1AC + 0.4GC + 0.4AE + 0.4EG + 0.2G + 0.4AJ + 0.7GJ + \\ & 0.1E \sim J + 0.1EE \sim \text{-----(4)} \end{aligned}$$

We get (4) as the solution, i.e., nine reducts which discern $Y=1$ from the other Ys.

The value of each term in (4) expresses the degree of discernibility.

Incidentally, (4) becomes the same as the five reducts of (1) if we limit the value to equal to or more than 0.4.

The case of categorical data with grades, differs from the case of ordinary categorical data and the category of each reduct implies an original category with a grade. The category of each reduct of (4) implies the grade of each category of U1 in Table 2 because only U1 is classified to

$Y=1$.

For example, AC of reduct 0.1AC, i.e., the 1st term of (4), implies $0.7A * 0.8C$. It is therefore convenient to add the original grade of U to each reduct as follows.

$$\begin{aligned} & 0.1/0.7A * 0.8C + 0.4/0.9G * 0.8C + 0.4/0.7A * 0.6E + \\ & 0.4/0.6E * 0.9G + 0.2/0.9G + 0.4/0.7A * 1.0J + \\ & 0.7/0.9G * 1.0J + 0.1/0.6E \sim * 1.0J + 0.1/0.6E * 0.6E \sim \text{-----(5)} \end{aligned}$$

4.2. Knowledge expression in the case of a single object

The knowledge acquired from (5) is expressed as follows.

“ $Y=1$ is discerned by the degree of 0.1 at least if only 0.7A or more and 0.8C or more, or by the degree of 0.4 at least if only 0.9G or more and 0.8C or more, or -----, or by the degree of 0.1 at least if only 0.6E or less and 1.0J, or by the degree of 0.1 at least if only 0.6E or more and 0.6E or less, i.e., if only 0.6E”.

Applying this knowledge expression to the case of the real world described as an example in Chapter 2, we get the following. Here, only those reducts having relatively larger degrees are picked up.

“All of the cars can be discerned perfectly as sporty from not-sporty, (from the 7th term) by the degree of 0.7 at least if only the cars are very low in height and have very showy body colors, or (from the 2nd term) by the degree of 0.4 at least if only the cars are very low in height and have silhouettes with very strong trapezoidal degrees, or (from the 3rd term) by the degree of 0.4 at least if only the cars have fairly personality-shaped head lamps and have more or less big and conspicuous tires”.

4.3. Reduct calculation in the case of plural objects

From Table 2, the three Us of U2, U3, and U4 are classified to $Y=2$. (2) and (3) are applied to $Y=2$ in Table 5 of the discernibility matrix.

When U is more than one, an operation is to be executed for every U separately, because it is meaningless to execute an operation between two difference values based on two different Us.

The result of the operations for $Y=2$ form a set consisting of 11 reducts as follows.

$$\left. \begin{array}{l} \text{(from U2)} \quad 0.4A \sim + 0.1E + 0.7G \sim \\ \text{(from U3)} \quad 0.1A \sim + 0.4E \sim + 0.7G \sim + 1.0J \sim \\ \text{(from U4)} \quad 0.4C \sim + 0.4E \sim + 0.2G \sim + 1.0J \sim \end{array} \right\} \text{-----(6)}$$

Equally to (5), adding the original grade of U in Table 2 into (6), gives the following 11 reducts.

$$0.4/0.3A^{\sim} + 0.1/0.7E + 0.7/0.2G^{\sim} + 0.1/0.6A^{\sim} + 0.4/0.2E^{\sim} + 0.7/0.2G^{\sim} + 1.0/0.0J^{\sim} + 0.4/0.4C^{\sim} + 0.4/0.2E^{\sim} + 0.2/0.7G^{\sim} + 1.0/0.0J^{\sim} \text{ -----}(7)$$

4.4. Knowledge expression in the case of plural objects

When the objects number more than one, it is sometimes possible to briefly express them by unifying these reducts having the same categories, depending on the grades and degrees.

If $p \geq q$

$$s/pA + t/qA \text{ -----}(8)$$

(8) represents “by degree s if pA or more, or by degree t if qA or more”.

Here, if $p \geq q$ and $s \leq t$, then the two reducts of (8) can be unified as follows and can be expressed briefly as “by degree t if qA or more”

$$s/pA + t/qA = t/qA$$

However, when $p \geq q$ and $s > t$, the two reducts of (8) cannot be unified.

The above mentioned expression changes into the opposite expression for

$$s/pA^{\sim} + t/qA^{\sim} \text{ -----}(9)$$

Here, if $p \geq q$ and $s \geq t$, then the two reducts of (9) can be unified as follows and can be expressed as “by degree s if pA or less”.

$$s/pA^{\sim} + t/qA^{\sim} = s/pA^{\sim}$$

However, when $p \geq q$ and $s < t$, the two reducts of (9) cannot be unified.

Figure 1 shows the foregoing description.

Figure 1: Relationship between the grade and degree in the same category

If the degree is not called into account, when the above-mentioned $p \geq q$, (8) is expressed simply as “if qA or more”, because the comparison between s and t is ignored.

The knowledge expression of (7) is, when picking up only some reducts, as follows.

“(from the 1st and 4th terms) $Y=2$ is discerned by the degree 0.1 at least if only 0.6A or less, and moreover, by the degree 0.4 at least if 0.3A or less”.

But if the degree is not called into account, the expression is simply “ $Y=2$ is discerned if only 0.6A or less”.

Then, “(from the 2nd and 5th terms) $Y=2$ is discerned by the degree 0.1 at least if only 0.7E or more, or by the degree 0.4 at least if only 0.2E or less”. But if the degree is not called into account, the expression is simply “ $Y=2$ is discerned if only 0.7E or more, or 0.2E or less”.

Not calling the degree into account means to accept all of the degrees. For the purpose of practical use, however, it is better to set a threshold and to cut off everything below the threshold and take everything above the threshold. In this way, the unifying increases and the reducts become easy to understand.

5. REDUCT CALCULATION FOR LARGE SIZE DATA

When data is of a large size and its discernibility matrix has plural U s in both the rows and columns as shown in Table 8, reducts can be found as well through the same process as described in the foregoing chapter.

Table 8 Discernibility matrix of large size data

G_{nm} is a set of differences given by the categories of U_n different from U_m

U	n+1	n+2	...	m
1	$G_1(n+1)$	$G_1(n+2)$	...	G_{1m}
2	$G_2(n+1)$	$G_2(n+2)$	...	G_{2m}
...	...	...	...	...
n	$G_n(n+1)$	$G_n(n+2)$	...	G_{nm}

The reducts for Table 8 are calculated by applying the

process described in chapter 4 to each of the following.

(from U1) $G1(n+1) * G1(n+2) * \dots * G1m$
 (from U2) $G2(n+1) * G2(n+2) * \dots * G2m$
 (...) $\dots * \dots * \dots * \dots$

6. CONCLUSION

This paper proposed a method of calculating the minimal reduct of the rough set theory for data such that the degree belonging to a certain category is given by a grade value $[0, 1]$.

In consideration of practical use, it is better to ignore reducts having small degrees.

In the real world, when attributes consist of *kansei* words in particular, grades are vague and the differences between grades, e.g., between grade 0.6 and 0.7, are almost meaningless. Reducts with a degree of 0.1 or so are also meaningless. It is perhaps better to induce some threshold as mentioned in chapter 4.

REFERENCES

1. S. Yokoyama, S. Tsumoto, K. Matsuoka, M. Fukunami; Data Mining in Laboratory Data; Proceedings of 16th Fuzzy System Symposium, 475-478 (2000).
2. N. Mori, R. Takanashi, T. Kumeda, K. Kumamaru; Proposal of Attribute Reduct as the Condition of Making a few new Products meet Diversified Preferences; Proceedings of 16th Fuzzy System Symposium, 305-306 (2000).
3. R. Takanashi, N. Mori, J. Sanui, M. Tian; Reduct Calculation and Consideration of Large Scale Data—Reduct in Rough Set Theory for Large Scale Data (2); Proceedings of the 2nd Annual Conference of JSKE 2000, 49 (2000).
4. N. Mori, R. Takanashi; Design Support System based on Reasoning with the Conception of Rough Set Theory; Bulletin of Faculty of Arts, Tokyo Institute of Polytechnics, 3, 35-38 (1997).
5. N. Mori, R. Takanashi; A Consideration of Thinking Process in *Kansei*-judgment for Products; Proceedings of the 1st Annual Conference of JSKE 1999, 48 (1999).
6. K. Inoue, M. Hirokawa; Analysis of the Relationship between Cognitive Parts and Evaluation Words; Proceedings of the 2nd Annual Conference of JSKE 2000, 62 (2000).
7. A. Nakamura, S. Tsumoto; Rough Set and Expert System -- Rough Set Theory and Application; Suurikagaku (in Japanese) 378, 76-83 (1994).
8. N. Mori, R. Takanashi; Design Support System based on Reasoning with the Conception of Rough Set Theory; Bulletin of Faculty of Arts, Tokyo Institute of Polytechnics, 3, 36 (1997).
9. *ibid.*, 37.