

# Proposal for a Statistical Reduct Method for Decision Tables

Yuichi Kato<sup>1</sup>(✉), Tetsuro Saeki<sup>2</sup>, and Shoutarou Mizuno<sup>1</sup>

<sup>1</sup> Shimane University, 1060 Nishikawatsu-cho, Matsue City, Shimane 690-8504, Japan  
ykato@cis.shimane-u.ac.jp

<sup>2</sup> Yamaguchi University, 2-16-1 Tokiwadai, Ube City, Yamaguchi 755-8611, Japan  
tsaeki@yamaguchi-u.ac.jp

**Abstract.** Rough Sets theory is widely used as a method for estimating and/or inducing the knowledge structure of if-then rules from a decision table after a reduct of the table. The concept of the reduct is that of constructing the decision table by necessary and sufficient condition attributes to induce the rules. This paper retests the reduct by the conventional methods by the use of simulation datasets after summarizing the reduct briefly and points out several problems of their methods. Then a new reduct method based on a statistical viewpoint is proposed. The validity and usefulness of the method is confirmed by applying it to the simulation datasets and a UCI dataset. Particularly, this paper shows a statistical local reduct method, very useful for estimating if-then rules hidden behind the decision table of interest.

## 1 Introduction

Rough Sets theory was introduced by Pawlak [1] and used for inducing if-then rules from a dataset called the decision table. The induced if-then rules simply and clearly express the structure of rating and/or knowledge hiding behind the decision table. Such rule induction methods are needed for disease diagnosis systems, discrimination problems, decision problems and other aspects. Each data in the decision table is a sample data consisting of the tuple of condition attributes' values and the decision attribute value, and the first step for the rule induction is to find the condition attributes which do not have any relationship with the decision attribute, to remove them and finally to reduce the table. The use of those processes to obtain the reduced table is called a reduct. The conventional rough sets theory to induce if-then rules is based on the indiscernibility of the samples of the given table by their attributes, and a reduct by the conventional method also uses the same concept and various types of indiscernibility, their methods and algorithms for the reducts are proposed to date [2–6].

This paper retests the conventional reduct methods through the use of a simulation dataset and points out their problems and how they lack receptivity for datasets containing conflicting and/or indifferent data, such as real-world datasets, after summarizing the conventional rough sets and reduct methods.

**Table 1.** Example of a decision table

$U$	$(C(1)C(2)C(3)C(4)C(5)C(6))$	$D$
1	563242	3
2	256124	6
3	116226	1
4	416646	6
...	.....	...
$N - 1$	151252	2
$N$	513135	4

Then a new reduct method is proposed to overcome their problems from a statistical point of view. Specifically, the new method recognizes each sample data in the decision table as the outcomes of random variables of the tuple of the condition attributes and the decision attribute since the dataset is obtained from their population of interest. Accordingly, the problem of finding the reduct through the concept of discernibility can be replaced by the problem of finding the condition attributes which are statistically independent of the decision attribute. This paper proposes two reduct methods: one is called a global reduct which examines the statistical independence between each condition attribute and the decision attribute, and the other is called a local reduct which finds each condition attribute statistically independent of a decision attribute value. The validity and usefulness of both reduct methods are confirmed by applying them to the simulation dataset as well as a UCI dataset [7] prepared for machine learning. The local reduct particularly shows that it gives important clues in the search for the if-then rules hidden behind the decision table of interest.

## 2 Conventional Rough Sets and Reduct Method

Rough Sets theory is used for inducing if-then rules from a decision table  $S$ .  $S$  is conventionally denoted  $S = (U, A = C \cup \{D\}, V, \rho)$ . Here,  $U = \{u(i) | i = 1, \dots, |U| = N\}$  is a sample set,  $A$  is an attribute set,  $C = \{C(j) | j = 1, \dots, |C|\}$  is a condition attribute set,  $C(j)$  is a member of  $C$  and a condition attribute, and  $D$  is a decision attribute.  $V$  is a set of attribute values denoted by  $V = \bigcup_{a \in A} V_a$  and is characterized by an information function  $\rho: U \times A \rightarrow V$ . Table 1 example where  $|C| = 6$ ,  $|V_{a=C(j)}| = M_{C(j)} = 6$ ,  $|V_{a=D}| = M_D = 6$ ,  $\rho(x = u(1), a = C(1)) = 5$ ,  $\rho(x = u(2), a = C(2)) = 5$ , and so on.

Rough Sets theory focuses on the following equivalence relation and equivalence set of indiscernibility:  $I_C = \{(u(i), u(j)) \in U^2 | \rho(u(i), a) = \rho(u(j), a), \forall a \in C\}$ .  $I_C$  is an equivalence relation in  $U$  and derives the quotient set  $U/I_C = \{[u_i]_C | i = 1, 2, \dots\}$ . Here,  $[u_i]_C = \{u(j) \in U | (u(j), u_i) \in I_C, u_i \in U\}$ .  $[u_i]_C$  is an equivalence set with the representative element  $u_i$  and is called an element set of  $C$  in Rough Sets theory [2]. Let be  $\forall X \subseteq U$  then  $X$  can be approximated like  $C_*(X) \subseteq X \subseteq C^*(X)$  by use of the element set. Here,

$$C_*(X) = \{u_i \in U \mid [u_i]_C \subseteq X\}, \quad (1)$$

$$C^*(X) = \{u_i \in U \mid [u_i]_C \cap X \neq \emptyset\}, \quad (2)$$

$C_*(X)$  and  $C^*(X)$  are called the lower and upper approximations of  $X$  by  $C$  respectively. The pair of  $(C_*(X), C^*(X))$  is usually called a rough set of  $X$  by  $C$ . Specifically, let  $X = D_d = \{u(i) \mid (\rho(u(i), D) = d)\}$  then  $C_*(X)$  is surely a set satisfying  $D = d$  and  $C^*(X)$  is possibly, and they derive if-then rules of  $D = d$  with necessity and possibility, respectively.

The conventional Rough Sets theory seeks a minimal subset of  $C$  denoted with  $B(\subseteq C)$  satisfying the following two conditions:

- (i)  $B_*(D_d) = C_*(D_d)$ ,  $d = 1, 2, \dots, M_D$ .
- (ii)  $a(\in B)$  satisfying  $(B - \{a\})_*(D_d) = C_*(D_d)$  ( $d = 1, 2, \dots, M_D$ ) does not exist.

$B(\subseteq C)$  is called a relative reduct of  $\{D_d \mid d = 1, \dots, M_D\}$  preserving the lower approximation, and is useful for finding if-then rules with necessity, since redundant condition attributes have been already removed from  $C$ . In the same way, a relative reduct preserving the upper approximation can be also defined and obtained.

LEM1 algorithm [2] and the discernibility matrix method (DMM) [3] are well known as representative ways to perform reducts. Figure 1 shows an example of LEM1, and  $A$  and  $\{d\}^*$  at Line 1 of the figure respectively correspond to  $C$  and  $\{D_d \mid d = 1, \dots, M_D\}$  in this paper. LEM1, from Line 6 to 15 in the figure, in principle, checks and executes (i) and (ii) for all the combinations of the condition attributes.

DMM [3] at first forms a symmetric  $N \times N$  matrix having the following  $(i, j)$  element  $\delta_{ij}$ :

$$\begin{aligned} \delta_{ij} &= \{a \in C \mid \rho(u(i), a) \neq \rho(u(j), a)\}; \\ &\exists d \in D, \rho(u(i), d) \neq \rho(u(j), d) \text{ and } \{u(i), u(j)\} \cap Pos(D) \neq \emptyset, \\ &= *; \quad \textit{otherwise}(U - Pos(D)). \end{aligned}$$

Here,  $Pos(D) = \cup_{d=1}^{M_D} C_*(D_d)$  and  $*$  denotes ‘‘don’t care’’. Then, a relative reduct preserving the lower approximation can be obtained by the following expression:

$$F^{reduct} = \wedge_{i,j:i < j} \delta_{ij}. \quad (3)$$

Here,  $F^{reduct}$  is called a discernible function and indicates that reducts are obtained by arranging all of respective discernibility.

### 3 Retests of the Conventional Reduct Method

Here we retest the ability of the reducts obtained through LEM1 and DMM by use of a simulation dataset. The literatures [8–11] show ways of how to generate simulation datasets. Specifically, let (a) generate the condition attribute values of  $u(i)$ , that is,  $u^C(i) = (v_{C(1)}(i), v_{C(2)}(i), \dots, v_{C(|C|)}(i))$  by the use of random numbers with a uniform distribution and (b) determine the decision attribute

Line No.	Algorithm to compute a single global covering
1	(input: the set $A$ of all attributes, partition $\{d\}^*$ on $U$ ; output: a single global covering $R$ );
2	Begin
3	compute partition $A^*$ ;
4	$P := A$ ;
5	$R := \emptyset$ ;
6	<b>if</b> $A^* \leq \{d\}^*$
7	<b>Then</b>
8	<b>Begin</b>
9	<b>for</b> each attribute $a$ in $A$ <b>do</b>
10	<b>Begin</b>
11	$Q := P - \{a\}$ ;
12	compute partition $Q^*$ ;
13	<b>if</b> $Q^* \leq \{d\}^*$ <b>then</b> $P := Q$
14	<b>end</b> {for}
15	<b>end</b> {then}
16	<b>end</b> {algorithm}

Fig. 1. Example of LEM1 algorithm

value of  $u(i)$ , that is,  $u^D(i)$  by use of if-then rules specified in advance and in the hypotheses shown in Table 2, and repeat the (a) and (b) processes by  $N$  times. Table 1 shows an example dataset generated by the use of those procedures with the following if-then rule  $R(d)$  specified in advance:

$$R(d) : \text{if } Rd \text{ then } D = d \quad (d = 1, \dots, M_D = 6), \tag{4}$$

where  $Rd = (C(1) = d) \wedge (C(2) = d) \vee (C(3) = d) \wedge (C(4) = d)$ . The results of retesting both methods using the  $N = 10000$  dataset showed  $F_{LEM1}^{reduct} = F_{DMM}^{reduct} = C(1) \wedge C(2) \wedge C(3) \wedge C(4) \wedge C(5) \wedge C(6)$  while the results were expected to be  $F^{reduct} = C(1) \wedge C(2) \wedge C(3) \wedge C(4)$  from the rules (4) specified in advance. The retest experiment was repeated three times by changing the generated dataset and obtained the same results, or in other words, the expected reducts were not obtained.

These results are clearly derived from the indiscernibility and/or discernibility caused by the element set in the conventional Rough Sets theory. The element sets couldn't distinguish the differences between samples by the if-then rules (see Hypothesis 1 in Table 2) or those obtained by chance (see Hypothesis 2 and 3 in Table 2). On the other hand, real-world datasets will contain all kinds of data generated by Hypotheses 1, 2 and 3. In other words, the conventional Rough Sets theory does not have any abilities adaptive to sample data caused by chance.

**Table 2.** Hypotheses with regard to the decision attribute value

Hypothesis 1	$u^C(i)$ coincides with $R(k)$ , and $u^D(i)$ is uniquely determined as $D = d(k)$ (uniquely determined data)
Hypothesis 2	$u^C(i)$ does not coincide with any $R(d)$ , and $u^D(i)$ can only be determined randomly (indifferent data)
Hypothesis 3	$u^C(i)$ coincides with several $R(d)$ ( $d = d1, d2, \dots$ ), and their outputs of $u^C(i)$ conflict with each other. Accordingly, the output of $u^C(i)$ must be randomly determined from the conflicted outputs (conflicted data)

## 4 Proposal of Statistical Reduct

### 4.1 Statistical Global Reduct Method

As mentioned in Sect. 3, conventional reduct methods are unable to reproduce the forms of reducts specified in advance from the decision table due to a lack of abilities adaptive to the indifferent and conflicted samples in datasets, despite the fact that real-world datasets will have such samples. This paper studies this problem with reducts from the view of STRIM (Statistical Test Rule Induction Method) [8–11]. STRIM regards the decision table as a sample set obtained from the population of interest. According to a statistical model,  $u(i) = (v_{C(1)}(i), v_{C(2)}(i), \dots, v_{C(|C|)}(i), u^D(i))$  is an outcome of the random variables of  $A = (C(1), C(2), \dots, C(|C|), D) = (C, D)$  (hereafter, the names of the attributes are also used as the random variables). Next, the following probability model will be specified: For any  $j$ ,  $P(C(j) = v_{C(j)}(k)) = p(j, k)$ ,  $\sum_{k=1}^{M_{C(j)}} p(j, k) = 1$ . For any  $j1 \neq j2$ ,  $P(C(j1) = v_{C(j1)}(k1), C(j2) = v_{C(j2)}(k2)) = p(j1, k1)p(j2, k2)$ , that is,  $C(j1)$  and  $C(j2)$  are independent of each other. According to the rules specified in (4), if  $C = (1, 1, 2, 3, 4, 5)$  (hereafter (112345) briefly), for example, then  $P(D = 1) = 1.0$  by use of Hypothesis 1 in Table 2. If  $C = (123456)$  then  $P(D = 1) = 1/M_D = 1/6$  by use of Hypothesis 2. If  $C = (112256)$  then  $P(D = 1) = 1/2$  by use of Hypothesis 3. Generally, the outcome of random variable  $D$  is determined by the outcome of  $C$ , if-then rules (generally unknown) and the hypothesis shown in Table 2. Consequently, the following expression is obtained:

$$P(D, C) = P(D|C)P(C = (v_{c(1)}(i), v_{c(2)}(i), \dots, v_{c(|C|)}(i))). \quad (5)$$

Here,  $P(D|C)$  is the condition probability of  $D$  by  $C$  and dependent on the if-then rules to be induced by the use of the dataset  $\{u(i) = (v_{C(1)}(i), v_{C(2)}(i), \dots, v_{C(|C|)}(i), u^D(i)) | i = 1, \dots, N\}$ .

In the special case, if  $C(j)$  does not exist in the condition part of the if-then rules to be induced, then  $D$  is independent of  $C(j)$ , that is  $P(D, C(j)) = P(D|C(j))P(C(j)) = P(D)P(C(j))$ . This independency between  $D$  and  $C(j)$  can be used for a reduct of the decision table. The problem of whether they are

independent or not can be easily dealt with using a statistical test of hypotheses by the use of  $\{u(i) = (v_{C(1)}(i), v_{C(2)}(i), \dots, v_{C(|C|)}(i), u^D(i)) | i = 1, \dots, N\}$ . Specifically, specification and testing of the following null hypothesis  $H0(j)$  and its alternative hypothesis  $H1(j)$  ( $j = 1, \dots, |C|$ ) were implemented:

$H0(j)$ :  $C(j)$  and  $D$  are independent of each other.

$H1(j)$ :  $C(j)$  and  $D$  are not independent of each other.

This paper adopts a Chi-square test since it is a standard method for testing the independency of two categorical variables by use of the contingency table  $M_{C(j)} \times M_D$ . Chi-square tests are widely used in a data mining area for feature selection [12]. The test statistic  $\chi^2$  of  $C(j)$  vs.  $D$  is

$$\chi^2 = \sum_{k=1}^{M_{C(j)}} \sum_{l=1}^{M_D} \frac{(f_{kl} - e_{kl})^2}{e_{kl}}, \tag{6}$$

where,  $f_{kl} = |U(C(j) = k) \cap U(D = l)|$ ,  $U(C(j) = k) = \{u(i) | v_{C(j)}(i) = k\}$ ,  $U(D = l) = \{u(i) | u^D(i) = l\}$ ,  $e_{kl} = n\hat{p}(j, k)\hat{p}(D, l)$ ,  $n = \sum_{k=1}^{M_{C(j)}} \sum_{l=1}^{M_D} f_{kl}$ ,  $\hat{p}(j, k) = f_{k\cdot}/n$ ,  $\hat{p}(D, l) = f_{\cdot l}/n$ ,  $f_{k\cdot} = \sum_{l=1}^{M_D} f_{kl}$ ,  $f_{\cdot l} = \sum_{k=1}^{M_{C(j)}} f_{kl}$ .  $\chi^2$  obeys a Chi-square distribution with degrees of freedom  $df = (M_{C(j)} - 1) \times (M_D - 1)$  under  $H0(j)$  and test condition [13]:  $n\hat{p}(j, k)\hat{p}(D, l) \geq 5$ . This paper proposes a reduct method to adopt only the  $C(j)$ s of  $H0(j)$  that were rejected and to construct a decision table composed by them since the test of the hypotheses can't control type II errors but only type I errors by a significance level. This paper names the proposed method the statistical global reduct (SGR) to distinguish it from the conventional method.

A simulation experiment was conducted to confirm the validity of the proposed method using the decision table of the samples of  $N = 10000$  used in Sect. 2, and the following procedures:

**Step1:** Randomly select samples by  $N_B = 3000$  from the decision table ( $N = 10000$ ), and form a new decision table;

**Step2:** Apply SGR to the new table, and calculate  $\chi^2$  every  $C(j)$ ;

**Step3:** Repeat Step1 and Step2  $N_r = 100$  times.

Table 3 shows the arrangement of the results of the simulation experiment, that is, the average (AVE), standard deviation (S.D.), maximum (Max) and minimum (Min) of  $\chi^2$ , and their corresponding  $p$ -values for every  $C(j)$  ( $j = 1, \dots, 6$ ). The table shows the following:

- (1) There are clear differences and big gaps in  $\chi^2$  values between those of  $C(1) - C(4)$  dependent on  $D$  and those of  $C(5)$  or  $C(6)$  independent of  $D$ , and are not overlapped by fluctuating ranges of the  $\chi^2$  value.
- (2) Accordingly, there are clear differences in  $p$ -values between them.
- (3) This reduct method of adopting rejected  $C(j)$ s of  $H0(j)$  and to construct a decision table composed by them is valid and useful since the notion reproduces the expected reduct.

**Table 3.** Results of test for independence by Bootstrap method ( $N = 10000$ ,  $N_B = 3000$ ,  $Nr = 100$ ):  $C(j)$  ( $j = 1, \dots, 6$ ) vs.  $D$

$\chi^2$		$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
AVE		341.69	360.59	313.27	327.40	24.79	25.23
S.D.		37.35	35.91	35.31	34.55	7.04	6.06
Max		451.19	471.25	412.19	418.39	47.83	40.61
Min		270.7	285.91	232.56	255.42	10.93	8.21
$p$ -value	Max	9.44E-80	6.84E-84	9.87E-72	5.28E-73	3.91E-03	2.52E-02
	Min	4.29E-43	3.98E-46	1.45E-35	4.59E-40	9.93E-01	9.99E-01

### 4.2 Statistical Local Reduct Method

In this section, a statistical local reduct (SLR) method which is applied for every decision attribute value, that is, every if-then rule, is proposed corresponding to the SGR method. The SLR method is applicable for the decision table which doesn't even have a relative reduct and/or the SGR. Specifically, (5) also holds at  $D = l$ , that is,

$$P(D = l, C) = P(D = l|C)P(C = (v_{c(1)}(i), v_{c(2)}(i), \dots, v_{c(|C|)}(i))). \quad (7)$$

Accordingly, the discussion in Sect. 4.1 holds even in the case where  $D$  is replaced in  $H0(j)$  and  $H1(j)$  with  $D = l$ , and the test of statistical independency between  $C(j)$  and  $D = l$  can be used for the reduct method for every  $D = l$  or if-then rule.

Table 4 shows a set of new rules specified in advance in order to confirm the validity and usefulness of the SGL and SLR method. For example, (1100001) at Rule 1 of the table means there is a rule that if  $(C(1) = 1) \wedge (C(2) = 1)$  then  $D = 1$ . The set of new rules doesn't include redundant condition attributes, that is,  $D$  is dependent on all condition attributes. The same simulation experiment in 4.1 was conducted replacing (4) with the set of rules in Table 4 and the results of the applied SGR method were summarized in Table 5 in the same way as Table 3. The validity of the SGR method is also confirmed since the table suggests the condition attributes seemed to be independent of  $D$  never existing, which coincides with the rules specified in advance. However, the structure of the if-then rules specified in advance is left almost unknown.

Then, the SLR method was applied for one of the same datasets of  $N_r = 100$  times used in Table 5. As one of examples, Table 6 shows the contingency table of the case of  $D = 1$  vs.  $C(j)$  ( $j = 1, \dots, 6$ ) and the results of a chi-square test of them with  $df = (M_{C(j)} - 1)$ , and suggests the following knowledge:

- (1) The  $p$ -values of  $C(5)$  and  $C(6)$  are quite high compared with those of the other condition attributes and indicate that  $C(5)$  and  $C(6)$  are independent of  $D = 1$ , that is, they are redundant and should be removed from the viewpoint of reduct.

**Table 4.** Example of if-then rules for a simulation experiment

Rule No.	specified rule: $(C(1)C(2), \dots, C(6)D)$
1	(1100001)
2	(0011001)
3	(0000222)
4	(2200002)
5	(0033003)
6	(0000333)
7	(4400004)
8	(0044004)
9	(0000555)
10	(5500005)
11	(0066006)
12	(0000666)

**Table 5.** Results of test for independence by Bootstrap method using sample dataset generated by rules in Table 4 ( $N = 10000$ ,  $N_B = 3000$ ,  $Nr = 100$ ):  $C(j)$  ( $j = 1, \dots, 6$ ) vs.  $D$

$\chi^2$		$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
AVE		209.56	206.35	244.57	234.94	252.45	219.59
S.D.		29.38	29.08	30.14	30.72	27.41	28.83
Max		293.56	290.93	318.17	317.52	310.48	314.13
Min		140.01	147.43	185.15	162.5	186.58	153.62
$p$ -value	Max	1.17E-47	3.95E-47	1.33E-52	1.80E-52	4.72E-51	8.69E-52
	Min	5.71E-18	2.51E-19	2.14E-26	4.03E-22	1.14E-26	1.81E-20

(2) The frequencies of  $C(1) = 1$ ,  $C(2) = 1$ ,  $C(3) = 1$  and  $C(4) = 1$  are relatively high compared with those of the rest of the same  $C(j)$  ( $j = 1, \dots, 4$ ) set. Accordingly, the combinations of  $C(j) = 1$  ( $j = 1, \dots, 4$ ) will most likely construct the rules of  $D = 1$ .

The above knowledge of (1) and (2) coincides with the specifications of Rules 1 and 2 in Table 4 and the same results also have been obtained from the cases of  $D = l$  ( $l = 2, \dots, 6$ ), and thus through them the validity and usefulness of the SLR method have been confirmed.

## 5 An Example Application on an Open Dataset

The literature [7] provides a lot of datasets for machine learning. This paper applied the SGR and SLR method for the “Car Evaluation” dataset included in



**Table 6.** Example of contingency table by local reducts ( $N = 3000, df = 5$ ):  $D = 1$  vs.  $C$

$D = 1$	$C$					
$ U(D = 1)  = 486$	1	2	3	4	5	6
1	<b>141</b>	<b>122</b>	<b>137</b>	<b>141</b>	78	88
2	64	67	89	80	95	74
3	82	86	59	82	74	79
4	58	77	53	52	93	101
5	79	58	83	75	71	64
6	62	76	65	56	75	80
$\chi^2$	52.83	35.01	66.42	53.16	2.99	13.33
$p$ -values	3.64E-10	1.5E-06	5.68E-13	5.00E-10	0.702	2.05E-2

**Table 7.** Arrangement of Car Evaluation dataset of UCI

unified attribute value	$C(1)$ : buying	$C(2)$ : maint	$C(3)$ : doors	$C(4)$ : person	$C(5)$ : lug boot	$C(6)$ : safety	$D$ : class (freq.)
1	vhigh	vhigh	2	2	small	low	unacc (1210)
2	high	high	3	4	med	med	acc (383)
3	med	med	4	more	big	high	good (69)
4	low	low	5more	–	–	–	vgood (65)

them. Table 7 shows the summaries and specifications of the dataset:  $|C| = 6, |V_{C(1)}| = |V_{C(2)}| = |V_{C(3)}| = 4, |V_{C(4)}| = |V_{C(5)}| = |V_{C(6)}| = 3, |V_D| = 4, N = |U| = |V_{C(1)}| \times |V_{C(2)}| \times |V_{C(3)}| \times |V_{C(4)}| \times |V_{C(5)}| \times |V_{C(6)}| = 1728$  which consists of every combination of condition attributes' values, and there were not any conflicted or identical samples. The frequencies of the decision attribute values extremely incline toward  $D = 1$  as shown in Table 7.

Table 8 shows the results obtained by the SGR method, that is,  $\chi^2$  and the corresponding  $p$ -values at every  $C(j)$  ( $j = 1, \dots, |C|$ ) and suggests that  $C(3)$  is independent of  $D$  and redundant compared to those of the other condition attributes, and should be removed from the viewpoint of the reduct.

**Table 8.** Results of global reduct for Car Evaluation dataset:  $D$  vs.  $C(j)$  ( $j = 1, \dots, 6$ )

		$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
$D$	$\chi^2$	189.24	142.94	10.38	371.34	53.28	479.32
	$p$ -value	5.92E-36	2.54E-26	3.20E-01	4.04E-77	1.03E-09	2.39E-100

**Table 9.** Results of local reduct for Car Evaluation dataset:  $D = \ell$  ( $\ell = 1, \dots, 4$ ) vs.  $C(j)$  ( $j = 1, \dots, 6$ )

		$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
$D = 1$	$\chi^2$	22.94	19.24	2.58	111.00	8.81	118.81
	$p$ -value	4.16E-05	2.44E-04	4.62E-01	7.88E-25	1.22E-02	1.59E-26
$D = 2$	$\chi^2$	11.77	10.77	3.19	192.56	6.52	194.25
	$p$ -value	8.21E-03	1.30E-02	3.64E-01	1.53E-42	3.85E-02	6.59E-43
$D = 3$	$\chi^2$	84.33	84.33	0.39	34.70	0.26	36.26
	$p$ -value	3.61E-18	3.61E-18	9.42E-01	2.92E-08	8.78E-01	1.34E-08
$D = 4$	$\chi^2$	70.20	28.60	4.23	33.08	37.69	130.00
	$p$ -value	3.87E-15	2.72E-06	2.38E-01	6.57E-08	6.53E-09	5.90E-29

**Table 10.** Examples of contingency table and  $\chi^2$  test by local reducts

(a) $D = 1$ vs. $C(j)$ ( $j = 1, \dots, 6$ )						
$V_{C(i)}$	$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
1	<b>360</b>	<b>360</b>	326	<b>576</b>	450	<b>576</b>
2	324	314	300	312	392	357
3	268	268	292	322	368	277
4	258	268	292	–	–	–
$\chi^2$	22.94	19.24	2.58	111.00	8.81	118.81
$p$ -value	4.16E-05	2.44E-04	4.62E-01	7.88E-25	1.22E-02	1.59E-26
(b) $D = 4$ vs. $C(j)$ ( $j = 1, \dots, 6$ )						
$V_{C(i)}$	$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
1	0	0	10	0	0	0
2	0	13	15	30	25	0
3	26	26	20	<b>35</b>	<b>40</b>	<b>65</b>
4	<b>39</b>	<b>26</b>	20	–	–	–
$\chi^2$	70.20	28.60	4.23	33.08	37.69	130.00
$p$ -value	3.87E-15	2.72E-06	2.38E-01	6.57E-08	6.53E-09	5.90E-29

Table 9 shows the results obtained by the SLR method, that is,  $\chi^2$  and the corresponding  $p$ -values at every  $D = l$  ( $l = 1, \dots, 4$ ) and  $C(j)$  ( $j = 1, \dots, |C|$ ) and suggests the following:

- (1)  $C(3)$  is commonly redundant at  $D = l$  ( $l = 1, \dots, 4$ ), which coincides with the results by the SGR method.
- (2) With regard to the if-then rule of  $D = 1$ ,  $C(5)$  is redundant. In the same way,  $C(2)$  and  $C(5)$  at  $D = 2$  are, and  $C(5)$  at  $D = 3$  is.

**Table 11.** Examples of estimated rules by use of the local reduct results in Table 10

Rule No.	examples of estimated rules ( $C(1)C(2)C(3)C(4)C(5)C(6)D$ )	distribution of decision values ( $n_1, n_2, n_3, n_4$ )	$p$ -value( $\chi^2$ )	accuracy	coverage
1	(0000011)	(576, 0, 0, 0)	3.58E-53(246.59)	1.00	0.476
2	(0001001)	(576, 0, 0, 0)	3.58E-53(246.59)	1.00	0.476
3	(0001011)	(192, 0, 0, 0)	1.03E-17(82.20)	1.00	0.157
4	(1000001)	(360, 72, 0, 0)	6.47E-11(50.43)	0.83	0.296
5	(0100001)	(360, 72, 0, 0)	6.47E-11(50.43)	0.83	0.296
6	(1000011)	(144, 0, 0, 0)	2.61E-13(61.64)	1.00	0.119
7	(0100011)	(144, 0, 0, 0)	2.61E-13(61.64)	1.00	0.119
8	(1001001)	(144, 0, 0, 0)	2.61E-13(61.64)	1.00	0.119
9	(0101001)	(144, 0, 0, 0)	2.61E-13(61.64)	1.00	0.119
10	(0000034)	(277, 204,30, 65)	2.24E-37(173.49)	0.113	1.000
11	(0000304)	(368, 144,24, 40)	1.24E-4(20.65)	0.069	0.615
12	(0000334)	(88, 64, 0, 40)	1.85E-39(183.14)	0.208	0.615
13	(4000034)	(52, 33, 20, 39)	1.24E-57(267.21)	0.27	0.600
14	(0400034)	(52, 46, 20, 26)	7.32E-31(143.30)	0.18	0.400
15	(4000334)	(16, 8, 0, 24)	–	0.50	0.369
16	(4403334)	(0, 0, 0, 4)	–	1.00	0.062

Table 10 shows examples of the contingency tables of  $D = 1$  (a) and  $D = 4$  (b), and their  $\chi^2$  by the SLR method, and suggests the following knowledge:

- (1) With regard to the if-then rules of  $D = 1$  (see (a)),the frequencies of  $C(1) = 1, C(2) = 1, C(4) = 1$  and  $C(6) = 1$  are distinctively high, which is statistically confirmed by their  $p$ -values. Accordingly, the if-then rules of  $D = 1$  are supposed to be constructed by the combinations of the set of  $\{C(1) = 1, C(2) = 1, C(4) = 1$  and  $C(6) = 1\}$  as shown in the knowledge obtained in 4.2.
- (2) In the same way, the if-then rules of  $D = 4$  are supposed to be constructed by the combinations of the set of  $\{C(1) = 4, C(2) = 4, C(4) = 3, C(5) = 3$  and  $C(6) = 3\}$  (see (b)).

Table 11 shows the trying rules (Rule 1-9 for  $D = 1$ , Rule 10-16 for  $D = 4$ ) based on the knowledge obtained above (1) and (2). Here Rule 15 and 16 do not have  $p$ -values since they do not satisfy the test condition  $n\hat{p}(j, k) \geq 5$  (for  $D = 4, n \geq \frac{5}{0.04} = 125$ ). Rule 1 and Rule 2 can be selected as one of the proper rules of  $D = 1$  since their  $p$ -values are extremely high and their indexes of accuracy and coverage are in good condition compared to the other trying rules. To express those rules by use of the original notation in Table 7, if person=2  $\vee$  safety=low

then  $\text{class}=\text{unacc}$ , is obtained with the  $\text{coverage}=(576+576-192)/1210 \approx 0.793$  and  $\text{accuracy}=1.0$ .

In the same way, the rule that if  $\text{buying}=\text{low} \wedge \text{safety}=\text{high}$  then  $\text{class}=\text{vgood}$  (Rule 13) is thought to be proper since the  $p$ -value is the best and the indexes of accuracy and coverage are moderate among the trying rules for  $D = 4$  although the number of sample data for  $D = 4$  is small compared to that for  $D = 1$ . Both rules estimated coincide with our common sense and the validity and usefulness of the SGR and SLR method have been experimentally confirmed through the use of an open dataset.

## 6 Conclusions

The Rough Sets theory has been used for inducing if-then rules from the decision table and clarifying the structure of rating and/or knowledge hidden behind the dataset. The first step in inducing the rules is to find reducts of the condition attributes. This paper retested the conventional reduct methods LEM1 [2] and DMM [3] by a simulation experiment after summarizing the conventional rough sets theory and pointed out their problems. Then, this paper proposed a new reduct method to overcome the problems of the conventional method from the view of STRIM [8–11]. Specifically, the SGR and SLR methods were proposed and their validity and usefulness were confirmed by a simulation experiment and application to an open dataset of UCI for machine learning. The SLR method should be recognized to be particularly useful for not only reducts of condition attributes but also inducing if-then rules.

## References

1. Pawlak, Z.: Rough sets. *Int. J. Inf. Comput. Sci.* **11**(5), 341–356 (1982)
2. Grzymala-Busse, J.W.: LERS – A system for learning from examples based on rough sets. In: Slowiński, R. (ed.) *Handbook of Applications and Advances of the Rough Sets Theory*. Theory and Decision Library, vol. 11, pp. 3–18. Kluwer Academic Publishers, Netherlands (1992)
3. Skowron, A., Rausser, C.M.: The Discernibility Matrix and Functions in Information Systems. In: Slowiński, R. (ed.) *Handbook of Application and Advances of Rough Sets Theory*. Theory and Decision Library, vol. 11, pp. 331–362. Kluwer Academic Publishers, Netherlands (1992)
4. Pawlak, Z.: Rough set fundamentals; KFIS Autumn Conference Tutorial, pp. 1–32 (1996)
5. Ślęzak, D.: Various approaches to reasoning with frequency based decision reducts: a survey. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds.) *Rough Set Method and Applications*, vol. 56, pp. 235–285. Physical-Verlag, Heidelberg (2000)
6. Bao, Y.G., Du, X.Y., Deng, M.G., Ishii, N.: An efficient method for computing all reducts. *Trans. Jpn. Soc. Artif. Intell.* **19**(3), 166–173 (2004)
7. Asunccion, A., Newman, D.J.: UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine (2007). <http://www.ics.edu/~mlearn/MLRepository.html>

8. Matsubayashi, T., Kato, Y., Saeki, T.: A new rule induction method from a decision table using a statistical test. In: Li, T., Nguyen, H.S., Wang, G., Grzymala-Busse, J., Janicki, R., Hassanien, A.E., Yu, H. (eds.) RSKT 2012. LNCS, vol. 7414, pp. 81–90. Springer, Heidelberg (2012)
9. Kato, Y., Saeki, T., Mizuno, S.: Studies on the necessary data size for rule induction by STRIM. In: Lingras, P., Wolski, M., Cornelis, C., Mitra, S., Wasilewski, P. (eds.) RSKT 2013. LNCS, vol. 8171, pp. 213–220. Springer, Heidelberg (2013)
10. Kato, Y., Saeki, T., Mizuno, S.: Considerations on rule induction procedures by STRIM and their relationship to VPRS. In: Kryszkiewicz, M., Cornelis, C., Ciucci, D., Medina-Moreno, J., Motoda, H., Raś, Z.W. (eds.) RSEISP 2014. LNCS, vol. 8537, pp. 198–208. Springer, Heidelberg (2014)
11. Kato, Y., Saeki, T., Mizuno, S.: Proposal of a statistical test rule induction method by use of the decision table. *Appl. Soft Comput.* **28**, 160–166 (2015). Elsevier
12. Kotu, V., Deshpande, B.: *Predictive Analytics and Data Mining*, 1st edn. Elsevier, Amsterdam (2014)
13. Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K.: *Probability and Statistics for Engineers and Scientists*, 8th edn. Pearson Prentice Hal, New Jersey (2007)