DATA MANAGEMENT OF FIELD INSPECTION FOR HIGHWAY BRIDGES BASED ON ADVANCED DATA MINING TECHNIQUE

A. MIYAMOTO [1*], and H. EMOTO[1†]

[1] *Graduate School of Science & Engineering, Yamaguchi University, Japan*

ABSTRACT

In this study, the acquisition of rule-type knowledge from field inspection data on highway bridges is enhanced by introducing an improvement to a traditional data mining technique, i.e. applying the rough set theory to the traditional decision table reduction method. The new rough set theory approach helps in cases of exceptional and contradictory data, which in the traditional decision table reduction method are simply removed from analyses. Instead of automatically removing all apparently contradictory data cases, the new method determines whether the data really is contradictory and therefore must be removed or not. The new method is tested with real data on bridge members including girders and filled joints in bridges owned and managed by a highway corporation in Japan. There are, however, numerous inconsistent data in field data. A new method is therefore proposed to solve the problem of data loss. The new method reveals some generally unrecognized decision rules in addition to generally accepted knowledge. Finally, a computer programs is developed to perform calculation routines, and some field inspection data on highway bridges is used to show the applicability of the proposed method.

Keywords: Data mining, Rough sets, Contradictory data, Bridge inspection, Highway bridge

1. INTRODUCTION

In this study, an attempt is made to acquire rule-type knowledge from large volumes of data stored in a field inspection database on bridges in expressway networks by a data mining method for efficient bridge maintenance. As a data mining method, the decision table reduction method based on the concept of the rough set theory (Pawlak 1982; Yokomori and Kobayashi 1994). Then, data with the same condition attributes but different decision attributes is considered contradictory data and is removed when rules are extracted for decision table reduction. Actual field inspection data, however, sometimes contains numerous contradictory data, so most of the field inspection data, although collected in large quantities, becomes useless.

To cope with the above problem, contradictory data that is removed when extracting rules, may be minimized by adding a function for saving the large majority of contradictory data (an algorithm for

* Corresponding author: Email: miya818@yamaguchi-u.ac.jp

† Presenter: Email: emoto@yamaguchi-u.ac.jp

rescuing contradictory data). In this study, a contradictory data rescuing algorithm was proposed that should be added to the rough set theory and its effectiveness was verified. Various studies were also made on the effects of the addition of the proposed algorithm on the extraction of rules.

2. **DATA MINING BASED ON ROUGH SET THEORY** (Rissanen 2007)

This chapter presents a flow of steps to extract rules through data mining based on the rough set theory, using a simple example (decision table). Table 1 lists examples of field inspection data that relates condition and decision attributes.

2.1. Reduction of decision table

In Table 1, universal set ***U*** is specified by {1, 2, 3, 4, 5, 6, 7}, the set of condition attributes ***C*** is specified by {ID, type of deck, type of girder, whether the main girder is straight or curved, girder classification}, and the set of decision attributes ***D*** is {whether damage is incurred or not }. The value that the attribute can take ***ρ*** is ***ρ***(ID=1, and (deck type)= reinforced concrete(RC) deck), ***ρ***(ID=1, and (girder type)= H-section girder or I-section girder Box-section girder or …), etc. A discernibility matrix that is obtained based on Table 1 is shown in Table 2. Asterisks{*} indicate the indiscernabile cases with the same sample number and contradictory cases where the condition

Table 1: Examples of condition (e.g. deck type) and decision (e.g. damaged or not) attribute data

ID	Damaged or not	Type of deck	Type of girder	Straight or curved main girder	Girder classification
1	Damaged	RC	H/I-section girder	Straight main girder	Steel girder
2	Damaged	RC	Box girder	Straight main girder	Pre-stressed concrete girder
3	Not damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
4	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
5	Not damaged	RC	H/I-section girder	Straight main girder	Steel girder
6	Not damaged	PC	Box girder	Straight main girder	Pre-stressed concrete girder
7	Damaged	RC	Box girder	Straight main girder	Pre-stressed concrete girder

Table 2: Discernibility matrix obtained based on the decision table

	1	2	3	4	5	6	7
1	*						
2	*	*					
3	Deck type, straight or curved main girder	Deck type, girder type, straight or curved main girder, girder classification	*				
4	*	*	*	*			
5	*	Deck type, girder type, girder classification	*	Deck type, straight or curved main girder	*		
6	Deck type, girder type, girder classification	Deck type	*	Deck type, girder type, straight or curved main girder, deck classification	*	*	
7	*	*	Deck type, girder type, straight or curved main girder	*	Girder type, girder classification	Deck type	*

attributes are the same but the decision attributes are different. Rows 1 and 3 in Table 1 with different decision attributes are compared to each other. Difference is found in condition attributes, namely the type of deck (RC deck, or steel deck) and whether the main girder is straight or curved (straight main girder, or curved main girder). Discerning is possible if there is a difference in one of the condition attributes. The set of elements in {1,3} in Table 2 is therefore expressed as {type of deck, whether the main girder is straight or curved}. Rows 1 and 5 are also compared to each other. Decision attributes are different while condition attributes are the same. Both cases are therefore considered contradictory.

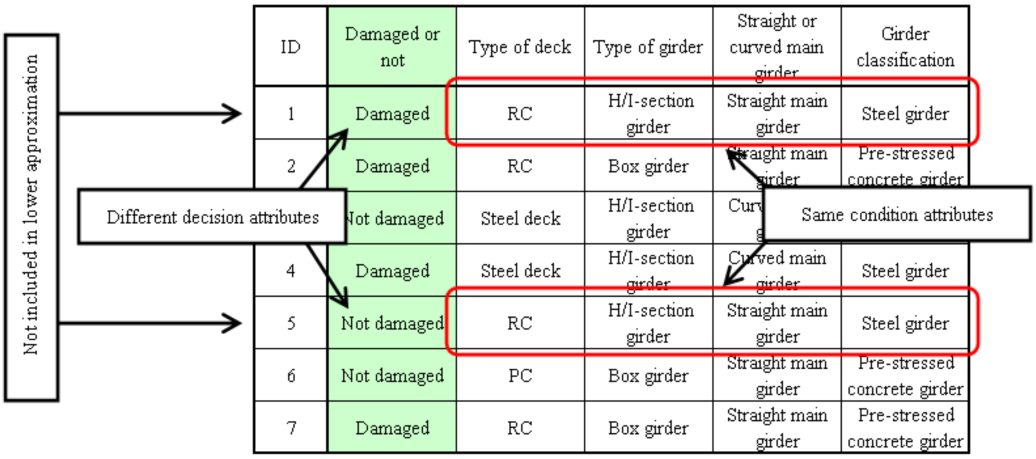
The elements in column 1 and row 5({1, 5}) are represented by {*} in Table 2. In this way, all the elements can be obtained in the discernibility matrix. Discernible conditions need to be met simultaneously, and therefore form the relation of logical disjunction. The decision table may therefore be reduced by calculating logical conjunctions for all the elements in Table 2. As an ultimate result of reduction, the combination of deck type and girder type and that of deck type and girder classification are conceivable. It is thus evident that one of these combinations is effective for discerning the decision attributes shown in Table 1.

2.2. Extraction of rules

The extraction of rules is explained in this section using an example of decision table shown in Table 1. First, decision attribute $\boldsymbol{D}$ is defined as whether damage is incurred or not. Then, the set for decision class D_1 (case where damage is incurred) is(=) {1, 2, 4, 7} and that for class D_2 (case where damage is not incurred) is(=) {3, 5, 6}. Then, $C_*(D_1)$ = {2,7} and $C_*(D_2)$ = {6}. It is meant that $C_*(D_1)$ is the lower approximation of decision class D_1. The attributes need to belong to decision class D_1. For {1,5} and {3,4}, decision attributes are different but all of the condition attributes are the same. Thus, these are contradictory data and not included in the lower approximation (Figure 1). Rules for determining that “damage is incurred” are extracted here as an example. A decision matrix obtained from Table 1 is shown in Table 3. “2” of $C_*(D_1)$, the lower approximation of decision class D_1, is compared with “3” of decision class D_2 for the attributes with different value from those in the decision table. The attributes for “2” are deck type: reinforced concrete, girder type: box girder, straight or curved main girder: straight main girder, and girder

Table 3: Decision matrix obtained based on the decision table

	3	5	6
2	{(deck type: reinforced concrete deck), (girder type: box girder), (straight or curved main girder: straight main girder), (girder classification: pre-stressed concrete girder)}	{(girder type: box girder), (girder classification: pre-stressed concrete girder)}	{(deck type: reinforced concrete deck)}
7	{(deck type: reinforced concrete deck), (girder type: box girder), (straight or curved main girder: straight main girder), (girder classification: pre-stressed concrete girder)}	{(girder type: box girder), (girder classification: pre-stressed concrete girder)}	{(deck type: reinforced concrete deck)}



ID	Damaged or not	Type of deck	Type of girder	Straight or curved main girder	Girder classification
1	Damaged	RC	H/I-section girder	Straight main girder	Steel girder
2	Damaged	RC	Box girder	Straight main girder	Pre-stressed concrete girder
[illegible]	Not damaged	Steel deck	H/I-section girder	Cur[illegible]	[illegible]
4	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
5	Not damaged	RC	H/I-section girder	Straight main girder	Steel girder
6	Not damaged	PC	Box girder	Straight main girder	Pre-stressed concrete girder
7	Damaged	RC	Box girder	Straight main girder	Pre-stressed concrete girder

Figure 1: Examples of cases not included in lower approximation

classification: prestressed concrete. If one of the attributes is different, discerning is possible. Thus, the relation is a logical disjunction. Similarly, "2" of $C_*(D1)$, the lower approximation of decision class D_1, is compared with "5" and "6" of decision class D_2. In order to discern "2" of $C_*(D_1)$, the lower approximation of decision class D_1,and decision class D_2, all of the conditions need to be met, which means a logical junction.

Calculations are also possible for "7" of $C_*(D_1)$, the lower approximation of decision class D_1. Finally, in order for all the conditions to be met in the decision matrix, each row needs to meet the condition for discerning. Each row is therefore a logical formula.

"2" in Table 3 is specified as:

{(deck type: RC deck) $\vee$ (girder type: box girder) $\vee$ (straight or curved main girder: straight main girder) $\vee$ (girder classification: prestressed concrete(PC) girder)} $\wedge$ {(girder type: box girder) $\vee$ (girder classification: PC girder)} $\wedge$ {(deck type: RC deck)}
= {(girder type: box girder) $\vee$ (girder classification: PC girder) $\wedge$ {(deck type: RC deck)} (1)

Similarly, "7" in Table 3 is specified as:

{(deck type: RC deck) $\vee$ (girder type: box girder) $\vee$ (straight or curved main girder: straight main girder) $\vee$ (girder classification: PC girder)} $\wedge$ {(girder type: box girder) $\vee$ (girder classification: PC girder)} $\wedge$ {(deck type: RC deck)}
= {(girder type: box girder) $\vee$ (girder classification: PC girder) $\wedge$ {(deck type: RC deck)} (2)

Applying logical formulas to the above produces the following:

{{(girder type: box girder) $\vee$ (girder classification: PC girder)} $\wedge$ {(deck type: RC deck)}} $\vee$ {{(girder type: box girder) $\vee$ (girder classification: PC girder) $\wedge$ {(deck type: RC deck)}}
= {(girder type: box girder) $\vee$ (girder classification: PC girder) $\wedge$ {(deck type: RC deck)}
= {(girder type: box girder) $\wedge$ (deck type: RC deck)} $\vee$ {(girder classification: PC girder) $\wedge$ (deck type: RC deck)} (3)

As a result, the following two rules are extracted.

(i) If the girder type is “box girder” and the deck type is “RC deck”, it is determined that “damage is incurred”.

(ii) If the girder classification is “PC girder” and the deck type is “RC deck”, it is determined that “dam-age is incurred”.

2.3. Indices for evaluating the reliability of extracted rules

The following indices are defined for determining the reliability of rules extracted by data mining.

2.3.1. Support index

This is an index representing the versatility of extracted rules. It is expressed as the percentage of cases that simultaneously satisfy the condition and decision parts of the extracted rule in all the cases, and obtained by

$$\text{Supp.} \quad (\Gamma|\Delta) = \frac{|\Gamma \wedge \Delta|}{n} \tag{4}$$

where, |Γ| and |Δ| are the numbers of data that satisfies logical formulas Γ and Δ, Δ is the extracted rule, |Γ ∧ Δ| is the number of data in the universal set U belonging to Δ, and n is the number of elements of the decision table.

2.3.2. Covering index (C.I.) value

The value of the covering index of the condition part of the extracted rule is expressed as the percentage of cases that comply with the extracted rule in the cases in the same decision class as for the decision of the extracted rule, and is obtained by

$$\text{Cov.} \quad (\Gamma|\Delta) = \frac{|\Gamma \wedge \Delta|}{|\Gamma|} \tag{5}$$

where, |Γ| and |Δ| are the numbers of data that satisfies logical formulas Γ and Δ. Δ is the extracted rule and |Γ ∧ Δ| is the number of data in the universal set *U* belonging to Δ.

3. PROPOSAL OF AN ALGORITHM FOR RESCUING CONTRADICTORY DATA

Described in this section are the problems encountered during the operation of rough sets using a decision table with numerous contradictory data. An algorithm for solving the problem is proposed.

3.1. Decision table with contradictory data and results

Contradictory data has the same condition attributes but different decision attributes. Contradictory data is removed during ordinary data mining. Data mining using the decision table with contradictory data shown in Table 4 (ID = {2, 6, 7, 8}, {3, 9, 10, 11}) produces the results shown in Table 5.

Table 4: Example of decision table with contradictory data

ID	Damaged or not	Type of deck	Type of girder	Straight or curved main girder	Girder classification
1	Damaged	RC	H/I-section girder	Straight main girder	Steel girder
2	Not damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
3	Not damaged	PC	Box girder	Curved main girder	Pre-stressed concrete girder
4	Damaged	RC	Box girder	Straight main girder	Pre-stressed concrete girder
5	Not damaged	RC	Box girder	Curved main girder	Pre-stressed concrete girder
6	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
7	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
8	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
9	Not damaged	PC	Box girder	Curved main girder	Pre-stressed concrete girder
10	Not damaged	PC	Box girder	Curved main girder	Pre-stressed concrete girder
11	Damaged	PC	Box girder	Curved main girder	Pre-stressed concrete girder

Table 5: C.I. values obtained by conventional methods

Rule for "damaged"	Deck type	Girder type	Straight or curved main girder	Girder classification	C.I.		1	4	6	7	8	11
Straight or curved main girder: Straight main girder			Straight main girder		2 / 6	0.333333	*	*				
Deck type: Reinforced concrete, and girder type: H/I-section girder	RC	H/I-section girder			1 / 6	0.166667	*					
Deck type: Reinforced concrete, and girder classification: steel girder	RC			Steel girder	1 / 6	0.166667	*					

In the rule that determines that damage is incurred, ID = {6, 7, 8, 11} is contradictory data and is re-moved from the cases for extracting rules. Contradictory data ID = {6, 7, 8, 11} is divided by decision attribute into majority contradictory data (ID = {6, 7, 8}; damaged) and minority contradictory data (ID = {2}; not damaged). If there are 101 contradictory data and the ratio of minority data to majority data is 1:100, all of the 101 data may be removed. Thus, contradictory data includes data that should be extracted for rules. As a means of solution, an algorithm for rescuing contradictory data is proposed below.

3.2. Basic concept of algorithm for rescuing contradictory data

The algorithm for rescuing contradictory data makes the data not applicable to the extracted rule as described in Section 3.1 applicable to the extracted rule. The processing procedure of the algorithm for rescuing contradictory data is shown below:

(i) Setting the percentage of data for extraction

The percentage of decision attributes in the contradictory data at which data is extracted is determined. The percentage should be higher than 50%.

(ii) Investigating contradictory data

Contradictory data is detected and the percentage of decision attributes is calculated.

(iii) Determining the data to be rescued

Contradictory data with the percentage of decision attributes exceeding the level specified in step (i) above is rescued.

(iv) Data mining

The decision table is reduced and rules are extracted using the data created in steps (i) through (iii).

3.3. Data mining by applying an algorithm for rescuing contradictory data

Data in Table 4 is subjected to data mining applying an algorithm for rescuing contradictory data. Then, the percentage of decision attributes for rule extraction is assumed to be 70%. First, contradictory data is investigated carefully. Contradictory data obtained from Table 4 is shown in Table 6.

ID= {2, 6, 7, 8} data in Table 6 shows that the percentage of decision attribute {Damaged or not = Damaged} is 75%, higher than a percentage of 70% for rule extraction. The majority, ID= {6, 7, 8}, is rescued and minority, ID= {2}, is removed. Similarly, the percentage of decision attribute {Damaged or not = Not damaged} in ID = {3, 9, 10, 11} is 75%, higher than a percentage of 70% for rule extraction. The majority, ID= {3, 9, 10}, is rescued and minority, ID= {11}, is removed. Thus, investigating contradictory data produces a decision table shown in Table 7. The values of covering index, which is an index of assessment of data mining results for determining that damage is incurred, is obtained from Table 7. The CI values are listed in Table 8.

ID = {6, 7, 8} shown in Table 4 is not applicable to extracted rules in Table 5, but is applicable to ex-tracted rules in Table 8 as a result of application of an algorithm for rescuing contradictory data.

The method proposed above is expected to enable effective use of contradictory data that has been de-leted because of inapplicability to extracted rules.

4. VERIFICATION OF EFFECTIVENESS OF THE ALGORITHM

This chapter verifies the effectiveness of an algorithm for rescuing contradictory data that has been added to data mining general-purpose software (Emoto et al. 2009).

Table 6: Example of investigation of contradictory data

ID	Condition attributes				Decision attributes		
	Deck type	Girder type	Straight or curved main girder	Girder classification	Damaged or not	Number of records	Percentage
2	Steel deck	H/I-section girder	Curved main girder	Steel girder	Not damaged	1	25%
6,7,8					Damaged	3	75%
3,9,10	PC	Box girder	Curved main girder	Pre-stressed concrete girder	Not damaged	3	75%
11					Damaged	1	25%

Table 7: Example of decision table after the investigation of contradictory data

ID	Damaged or not	Type of deck	Type of girder	Straight or curved main girder	Girder classification
1	Damaged	RC	H/I-section girder	Straight main girder	Steel girder
3	Not damaged	PC	Box girder	Curved main girder	Pre-stressed concrete girder
4	Damaged	RC	Box girder	Straight main girder	Pre-stressed concrete girder
5	Not damaged	RC	Box girder	Curved main girder	Pre-stressed concrete girder
6	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
7	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
8	Damaged	Steel deck	H/I-section girder	Curved main girder	Steel girder
9	Not damaged	PC	Box girder	Curved main girder	Pre-stressed concrete girder
10	Not damaged	PC	Box girder	Curved main girder	Pre-stressed concrete girder

Table 8: C.I. values in the case where an algorithm for rescuing contradictory data is applied

Rule for "damaged"	Deck type	Girder type	Straight or curved main girder	Girder classification	C.I.		1	4	6	7	8
Girder type: H/I-section girder		H/I-section girder			4 / 6	0.666667	*		*	*	*
Straight or curved main girder: Straight main girder			Straight main girder		2 / 6	0.333333	*	*			
Girder classification: Steel girder				Steel girder	4 / 6	0.666667	*		*	*	*
Deck type: Steel deck	Steel deck				3 / 6	0.5			*	*	*

For verification, inspection data collected on actual bridges were used. Data mining was done in the case where no algorithm for rescuing contradictory data was applied and in the case where an algorithm for rescuing contradictory data was applied where the rate of decision attribute for rule extraction was 70%. The results were compared in terms of support index and covering index values, indices for assessing the results of data mining described in Chapter 2. Then, the effect of the rate of decision attribute for rule extraction on results was also verified by applying an algorithm for rescuing contradictory data where the percentage of decision attribute for rule extraction was 60, 70 or 80%.

4.1. Bridge inspection data containing contradictory data for data mining

As bridge inspection data, a decision table obtained by preprocessing data on damage to the bearing to be analyzed (73 items), bearing property data (29 items), superstructure property data (57 items), expansion joint damage data (60 items) and expansion joint property data (29 items) was used. Preprocessing included data integration, and the concentration and clustering of condition attributes. Data was integrated by Ryosen Engineer's Co. Ltd., Hiroshima, Japan and engineering judgment was made by Hanshin Expressway Engineering Co. Ltd., Osaka, Japan for concentrating and clustering condition attributes.

Damage to bearings comes in several types. In this study, data mining was done for bridge seat concrete using the condition attributes shown in Table 9. Some of the bridge inspection data are listed in Table 10.

4.2. Verification of effectiveness of an algorithm for rescuing contradictory data

Data mining was done using bridge inspection data containing contradictory data prepared in Section 4.1 (decision table shown in Table 10) in the case where no algorithm for rescuing contradictory data was applied and in the case where an algorithm was applied at an extraction rate of 70%. Some of the results in the case where no algorithm was used are shown in Table 11. Some of the results in the case where an algorithm was applied at a percentage of decision at-tribute for rule extraction of 70% are shown in Table 12. Those results with high values of support index and covering index(C.I.) values, which are indices for assessing output results, were selectively presented.

Tables 11 and 12 were compared to each other. As a result, it was found that support and C.I. value increased considerably in Table 12 in the case where an algorithm was applied for rescuing contradictory data. Applying an algorithm for rescuing contradictory data enables deleted contradictory data to be rescued and makes the application of extraction rules possible.

Table 9: List of condition attributes to be selected for each type of bearing damage

Decision attributes			Condition attributes											
Type of bearing damage	Number of records		Bearing					Superstructure					Expansion joint	
	Damaged	Not damaged	Year completed	Stringer classification	Model	Fixed/ movable	Nominal tonnage	Girder classification	Deck type	Girder type	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder	Model	Damage
Bridge seat concrete	3582	5907	○	○	○	○	○	○	○	○	○	○		

Table 10: Some of the bridge inspection data (damage to bridge seat concrete)

Index	Bearing is damaged or not	Year bearing was completed	Classification of bearing stringer	Model of bearing	Bearing is fixed or movable	Nominal tonnage of bearing	Classification of superstructure girder	Superstructure deck type	Superstructure girder type	Bachi (plectrum)-shaped/skew bridge	Straight or curved superstructure main girder
1	Bridge seat concrete	A	Simple girder	Bearing plate support	Fixed	A	Steel girder	RC	H/I-section	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder
2	Bridge seat concrete	A	Simple girder	Bearing plate support	Fixed	A	Steel girder	RC	H/I-section	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder
3	Bridge seat concrete	A	Simple girder	Bearing plate support	Fixed	A	Steel girder	RC	H/I-section	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder
9487	Not damaged	E	Midpoint in continuous girder	Other	Movable	A	Steel girder	Steel deck	Box girder	None	Straight main girder
9488	Not damaged	E	Midpoint in continuous girder	Other	Movable	A	Steel girder	Steel deck	Box girder	None	Straight main girder
9489	Not damaged	E	Midpoint in continuous girder	Other	Movable	A	Steel girder	Steel deck	Box girder	None	Straight main girder

Table 11: Extracted rules with high support value for "bridge seat concrete" (algorithm not applied)

Rule for "bearing is damaged or not":" Bridge seat concrete"

No.	Bearing					Superstructure					Support	CI value	
	Year of completion	Stringer classification	Model	Fixed/ movable	Nominal tonnage	Girder classification	Deck type	Girder type	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder			
1	A							H/I-section	Bachi (plectrum)-shaped bridge	Curved main girder	0.007482	71 / 3582	0.019821
2	A					Steel girder			Skew bridge		0.005796	55 / 3582	0.015355
3	A						RC		Skew bridge		0.005796	55 / 3582	0.015355
4	A		Bearing plate support						Skew bridge		0.005585	53 / 3582	0.014796
5			Bearing plate support						Skew bridge	Straight main girder	0.005375	51 / 3582	0.014238
6	A							Box girder	Bachi (plectrum)-shaped bridge	Straight main girder	0.005058	48 / 3582	0.013400
7	A		Bearing plate support	Fixed				H/I-section		Curved main girder	0.004742	45 / 3582	0.012563
8		Simple girder	Bearing plate support				RC	Box girder	Bachi (plectrum)-shaped bridge	Straight main girder	0.004742	45 / 3582	0.012563
9	A		Bearing plate support	Movable			RC	Box girder		Straight main girder	0.004742	45 / 3582	0.012563
10	A							Box girder	Skew bridge		0.004637	44 / 3582	0.012284

Table 12: Extracted rules with high support value for "bridge seat concrete" (algorithm applied at a percentage of decision attribute for rule extraction of 70%)

Rule for "bearing is damaged or not":" Bridge seat concrete"

No.	Bearing					Superstructure					Support	CI value	
	Year of completion	Stringer classification	Model	Fixed/ movable	Nominal tonnage	Girder classification	Deck type	Girder type	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder			
1	A		Bearing plate support				RC			Straight main girder	0.112876	1016 / 3582	0.283640
2	A	Simple girder	Bearing plate support	Fixed						Straight main girder	0.086657	780 / 3582	0.217755
3	A		Bearing plate support	Fixed		Steel girder				Straight main girder	0.086324	777 / 3582	0.216918
4	A		Bearing plate support				RC		None		0.061215	551 / 3582	0.153825
5	A	Simple girder	Bearing plate support	Fixed					Bachi (plectrum)-shaped bridge		0.048328	435 / 3582	0.121441
6	A		Bearing plate support	Fixed				H/I-section			0.047884	431 / 3582	0.120324
7	A		Bearing plate support	Fixed			RC		Bachi (plectrum)-shaped bridge		0.047772	430 / 3582	0.120045
8	A	Simple girder	Bearing plate support	Fixed					None		0.046328	417 / 3582	0.116415
9	A		Bearing plate support	Fixed		Steel girder			None		0.046106	415 / 3582	0.115857
10	A		Bearing plate support					H/I-section	Bachi (plectrum)-shaped bridge		0.041551	374 / 3582	0.104411

4.3. Verification of effect of the percentage of decision attribute for rule extraction on results

Next, the effect that the percentage for rule extraction had on results was verified. Some of the results in the case where an algorithm for rescuing contradictory data was applied at a percentage for rule ex-traction of 60, 70 or 80% are shown in Tables 13, 14 and 15, respectively. The results in Tables 13 through 15 were compared with one another. It is evident that support and C.I. values,

Table 13: Results at a percentage for rule extraction of 60%

Rule for "bearing is damaged or not":" Bridge seat concrete"

No.	Bearing					Superstructure					Support	CI value	
	Year of completion	Stringer classification	Model	Fixed/ movable	Nominal tonnage	Girder classification	Deck type	Girder type	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder			
1	A		Bearing plate support				RC			Straight main girder	0.116755	1016 / 3582	0.283640
2	A	Simple girder	Bearing plate support	Fixed						Straight main girder	0.089635	780 / 3582	0.217755
3	A		Bearing plate support	Fixed		Steel girder				Straight main girder	0.089290	777 / 3582	0.216918
4	A		Bearing plate support				RC	H/I-section			0.075615	658 / 3582	0.183696
5	A		Bearing plate support				RC		None		0.063319	551 / 3582	0.153825
6	A	Simple girder	Bearing plate support	Fixed					Bachi (plectrum)-shaped bridge		0.049989	435 / 3582	0.121441
7	A		Bearing plate support	Fixed				H/I-section			0.049529	431 / 3582	0.120324
8	A		Bearing plate support	Fixed			RC		Bachi (plectrum)-shaped bridge		0.049414	430 / 3582	0.120045
9	A	Simple girder	Bearing plate support	Fixed					None		0.047920	417 / 3582	0.116415
10	A		Bearing plate support	Fixed		Steel girder			None		0.047690	415 / 3582	0.115857

Table 14: Results at a percentage for rule extraction of 70%

Rule for "bearing is damaged or not":" Bridge seat concrete"

No.	Bearing					Superstructure					Support	CI value	
	Year of completion	Stringer classification	Model	Fixed/ movable	Nominal tonnage	Girder classification	Deck type	Girder type	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder			
1	A		Bearing plate support				RC			Straight main girder	0.112876	1016 / 3582	0.283640
2	A	Simple girder	Bearing plate support	Fixed						Straight main girder	0.086657	780 / 3582	0.217755
3	A		Bearing plate support	Fixed		Steel girder				Straight main girder	0.086324	777 / 3582	0.216918
4	A		Bearing plate support				RC		None		0.061215	551 / 3582	0.153825
5	A	Simple girder	Bearing plate support	Fixed					Bachi (plectrum)-shaped bridge		0.048328	435 / 3582	0.121441
6	A		Bearing plate support	Fixed				H/I-section			0.047884	431 / 3582	0.120324
7	A		Bearing plate support	Fixed			RC		Bachi (plectrum)-shaped bridge		0.047772	430 / 3582	0.120045
8	A	Simple girder	Bearing plate support	Fixed					None		0.046328	417 / 3582	0.116415
9	A		Bearing plate support	Fixed		Steel girder			None		0.046106	415 / 3582	0.115857
10	A		Bearing plate support					H/I-section	Bachi (plectrum)-shaped bridge		0.041551	374 / 3582	0.104411

assessment indices, increased as the percentage for rule extraction decreased. This suggests that more data was rescued at a higher percentage for rule extraction. Thus, lowering the percentage for rule extraction enables data mining with more data. At a lower percentage for rule extraction, however, more ambiguous data is rescued, so extracted rules are more likely to be false. Then, it is necessary to analyze the quantity of contradictory data contained in extracted rules and to make engineering judgment of the content of the extracted rule.

Table 15: Results at a percentage for rule extraction of 80%

Rule for "bearing is damaged or not":" Bridge seat concrete"

No.	Bearing					Superstructure					Support	CI value	
	Year of completion	Stringer classification	Model	Fixed/ movable	Nominal tonnage	Girder classification	Deck type	Girder type	Bachi (plectrum)-shaped/skew bridge	Straight/curved main girder			
1	A	Simple girder	Bearing plate support	Fixed						Straight main girder	0.084930	780 / 3582	0.217755
2	A		Bearing plate support	Fixed		Steel girder				Straight main girder	0.084604	777 / 3582	0.216918
3	A		Bearing plate support	Fixed			RC			Straight main girder	0.084604	777 / 3582	0.216918
4	A		Bearing plate support				RC	Box girder		Straight main girder	0.058907	541 / 3582	0.151033
5	A		Bearing plate support	Fixed				H/I-section			0.046929	431 / 3582	0.120324
6	A	Simple girder	Bearing plate support	Fixed					None		0.045405	417 / 3582	0.116415
7	A		Bearing plate support	Fixed		Steel girder			None		0.045187	415 / 3582	0.115857
8	A		Bearing plate support	Fixed			RC		None		0.045187	415 / 3582	0.115857
9	A		Bearing plate support				RC		Bachi (plectrum)-shaped bridge	Straight main girder	0.039416	362 / 3582	0.101061
10	A		Bearing plate support				RC	Box girder	None		0.034734	319 / 3582	0.089056

5. CONCLUSIONS

The results obtained from this study can be summarized as follows:

1) Rescuing deleted contradictory data was made possible by applying an algorithm for rescuing contradictory data to rough sets.

2) In the verification using data on actual bridges, contradictory data was rescued. Thus, the effectiveness of an algorithm for rescuing contradictory data was verified.

3) Lowering the percentage of decision attribute for rule extraction enables the rescue of contradictory data. Reducing the percentage to an extremely low level, however, means the extraction of rules based on ambiguous data and is likely to output false results.

4) When setting the percentage of decision attribute for rule extraction, the contradictory data contained in the data to be handled should be analyzed to verify the contents of contradictory data and the percentages of majority and minority contradictory data. Establishing an appropriate method for using the percentage for rule extraction is a future task.

REFERENCES

Emoto,H., Takahashi,J., and Miyamoto, A. (2009):"Development of general-purpose data mining software in which rough sets are applied", Research Center for Environmental Safety (RCES), Yamaguchi University, Lifetime Man-agement Book No. 7, 1-42.

Pawlak, Z. (1982). "Rough sets", International Journal of Computer and Information Sciences, 11(5), 541 – 356.

Rissanen, T,, Miyamoto, A., Tarighat, A.., & Ishida, J. (2007). "Knowledge Discovery in Bridge Data-base by Rough Set based Data Mining", Ad-vances in Fuzzy Sets and Systems, Pushpa Pub-lishing House, 2(2), 179-216.

Yokomori, T. and Kobayashi, S. (1994). "Rough Sets and Decision Making". Mathematical Sci-ences, No. 375, 76 – 83.