PAVEMENT MAINTENANCE AND REHABILITATION DECISIONS DERIVED BY ROUGH SET THEORY

Jia-Ruey Chang[1], Ching-Tsung Hung[2], Gwo-Hshiung Tzeng[3], and Wen-Hsiung Hsiao[4]

ABSTRACT

Rough Set Theory (RST) is an induction based decision-making technique, which can extract useful information from attribute-value (decision) table. This study introduces RST into pavement management system (PMS) for maintenance and rehabilitation (M&R) strategy induction. An empirical study is conducted by using the pavement distress survey data collected by experienced pavement engineers of Taiwan Highway Bureau (THB) in 1999. Eighteen distress types and their corresponding M&R treatments were surveyed and recorded from seven county roads to establish the analytical database. The database consists of 2,348 records (2,115 records for rule induction, and 233 records for rule verification). On the basis of the verification results, total accuracy for the induced rules is as high as 93.6%, which illustrates that RST certainly can assist pavement engineers to easily remove redundant records, reduce attributes, discover association, and induce the most appropriate M&R strategy when they deal with the enormous pavement distress data.

KEY WORDS

pavement management systems (PMS), maintenance and rehabilitation (M&R), rough set theory (RST), decision-making, distress.

INTRODUCTION

Various distress types would occur to pavement because of dynamic loading, overweighed trucks, weak foundation, improper mix design, change of climates, etc (Huang 1993). Pavement distress survey requires the severity and coverage of considered distress types in order to adopt appropriate maintenance and rehabilitation (M&R) strategies. It is extremely important that if appropriate M&R strategies can be conducted at right time for specific distress. Appropriate M&R strategy can not only save long-term expense but keep the pavement above an acceptable serviceability. The objective of this study is to utilize RST technique in dealing with enormous pavement distress records to carry out redundant records

[1] Assistant Professor, Dept. of Civil Engineering, MingHsin University of Science & Technology, No.1 Hsin-Hsing Road, Hsin-Fong, Hsin-Chu, 304, Taiwan, Phone +886 3/559-3142 ext. 3295, FAX +886 3/557-3718, jrchang@must.edu.tw

[2] Ph.D. Candidate, Institute of Civil Engineering, National Central University, ChungLi, Taoyuan, 320, Taiwan, Phone +886 3/426-9270, FAX +886 3/422-7183, 92342008@cc.ncu.edu.tw

[3] Distinguished Chair Professor of Kainan University, No.1 Kainan Road, Luchu, Taoyuan County, 338, Taiwan, Phone +886 3/341-2500 ext. 1101, FAX +886 3/341-2430, ghtzeng@mail.knu.edu.tw

[4] Senior Specialist, Construction and Planning Agency, Ministry of Interior, No. 342, Sec. 2, Bar-Der Rd., Taipei, Taiwan, Phone +886 2/8771-2345, winters@cpami.gov.tw

removal, attributes reduction, association discovery, decision table simplification, and M&R strategies induction for improving the efficiency of M&R decision-making process.

ROUGH SET THEORY

Rough set, originally proposed by Pawlak (1982), is a mathematical tool used to deal with vagueness or uncertainty. Compared to fuzzy set, there are some advantages to rough set theory (Pawlak, Grzymala-Busse, Slowinski, and Ziarko 1995). One main advantage is that rough set does not need any pre-assumptions or preliminary information about the data, such as the grade of membership function in fuzzy set (Grzymala-Busse 1988). Recently, rough set theory and fuzzy set theory have been used to complement or incorporate (Radzikowska & Kerre 2002) each other rather than to compete (Dubois and Prade 1991). More detailed discussion about the process of rough set theory can refer to Walczak and Massart (1999). The original concept of approximation space in rough set can be described as follows.

Given an approximation space

$apr = (U, A)$

where U is the universe which is a finite and non-empty set, and A is the set of attributes. Then based on the approximation space, we can define the lower and upper approximations of a set.

Let X be a subset of U and the lower approximation of X in A is

$$\overline{apr}(A) = \{x \mid x \in U, U / Ind(A) \subset X\} \tag{1}$$

The upper approximation of X in A is

$$\underline{apr}(A) = \{x \mid x \in U, U / Ind(A) \cap X \neq \varnothing\} \tag{2}$$

where

$$U / Ind(A) = \{(x_i, x_j) \in U \cdot U,\ f(x_i, a) = f(x_j, a) \quad \forall a \in A\} \tag{3}$$

Eq. (1) represents the least composed set in A containing X, called the best upper approximation of X in A, and Eq. (2) represents the greatest composed set in A contained in X, called the best lower approximation.

After constructing upper and lower approximations, the boundary can be represented as

$$BN(A) = \overline{apr}(A) - \underline{apr}(A) \tag{4}$$

According to the approximation space, we can calculate reducts and decision rules. Given an information system $I = (U, A)$ then the reduct, $RED(B)$, is a minimal set of attributes $B \subseteq A$ such that $r_B(U) = r_A(U)$ where

$$r_B(U) = \frac{\sum card(\underline{B}X_i)}{card(U)} \tag{5}$$

denotes the quality of approximation of U by B.

Once the reducts have been derived, overlaying the reducts on the information system can induce the decision rules. A decision rule can be expressed as $\varnothing \Rightarrow \theta$, where $\varnothing$ denotes the conjunction of elementary conditions, $\Rightarrow$ denotes '*indicates*', and θ denotes the disjunction of elementary decisions.

The advantage of the induction based approaches (e.g. rough set and decision trees) is that it can provide the intelligible rules for decision-makers (DMs). These intelligible rules can help DMs to realize the contents of data sets.

ROUGH SET THEORY USED IN PAVEMENT MANAGEMENT PROCESS

Pavement condition collection plays an important role in pavement management process. The appropriate M&R strategies are decided according to correct condition information. Although the current decision support techniques can acquire valuable information from enormous data, the systematized analytic process is needed for the significant improvement of accuracy and efficiency. Systematized decision support system for pavement management must accomplish the following functions (Attoh-Okine 1997):

- Discover pavement management database to induce correct M&R rules
- Screen the redundant and useless data
- Simplify the database if too much useless information exists
- Describe the association between distress and M&R strategy as decision rules

According to literature reviews, data mining techniques actually applied to engineering issues are limited, especially RST application is extremely rare. RST provides a new direction for pavement knowledge discovery and decision table analysis. RST can be used in pavement database primarily to deduct attributes, remove redundant records, simplify decision table, and induce M&R rules. Attoh-Okine (1997) demonstrated the feasibility of RST to M&R decision induction and suggested more reality data for verification is required. Attoh-Okine (2002) further explored the potential applications of rough set theory and neural networks in concrete-faulting-performance modeling. The key characteristic of the proposed method is that the new decision table created by using the rough set analysis will free the neural network paradigm from redundancy.

EMPIRICAL STUDY: A CASE OF PAVEMENT M&R STRATEGY INDUCTION

PAVEMENT DISTRESS SURVEY AND M&R STRATEGY

Pavement distress survey is conducted for the purpose of monitoring the existing pavement condition and making the appropriate M&R activities. Generally, the severity and coverage should be separately identified and recorded for each distress type. For accurate, consistent, and repeatable distress surveys, one distress survey manual is required for clarifying the definition, severity, and coverage of each distress type. In the empirical study, pavement distress survey was carried out primarily following the Distress Identification Manual for the Long-Term Pavement Performance Program issued by FHWA. Furthermore, the adopted M&R strategy for each distress was identified according to the standardized M&R Guidance used in Taiwan Highway Bureau (THB).

PROBLEM DESCRIPTION

Pavement distress surveys were conducted on seven county roads in Chung-Li Engineering Section of THB by 8 experienced pavement engineers in 1999. The collected 2,348 records (2,115 records were randomly selected for rule induction; the rest of 233 records were for rule verification) are utilized in the empirical study. The seven county roads are 110, 112, 112A, 113, 113A, 114 and 115. Engineers were carrying surveys by walking or driving and recording the distress information and their corresponding M&R strategy. Then, the data are integrated as Table 1. The first column in Table 1 shows the record numbers, column 2 to column 19 illustrate the distress types, and the last column refers to the adopted M&R strategy. The details are described as the following:

Table 1: Analytic Database

Table: Modeling

2115 / 19	D1	D2	D3	...	...	...	...	...	...	...	...	...	...	...	...	D16	D17	D18	M&R
O:1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	1
O:2	1	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	1
O:3	2	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1
O:4	2	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1
O:5	2	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:6	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:7	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:8	2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:9	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:10	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:11	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:12	4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
O:13	5	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1
O:14	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1
O:15	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1
O:16	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:17	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:18	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:19	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:20	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:21	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:22	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:23	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:24	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:25	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:26	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:27	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:28	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:29	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:30	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:31	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:32	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:33	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:34	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:35	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:36	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:37	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:38	2	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:39	2	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:40	4	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:41	1	0	1	5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:42	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:43	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:44	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
O:45	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1

- This study explores 18 common distress types in Taiwan, which are represented from D1 to D18: D1. Alligator Cracking, D2. Block Cracking, D3. Longitudinal Cracking, D4. Transverse Cracking, D5. Edge Cracking, D6. Reflection Cracking, D7. Pothole, D8. Bleeding , D9. Rutting, D10. Corrugation, D11. Lane/Shoulder Drop-off, D12. Depression, D13. Structure Drop-off, D14. Utility Cut Patching, D15. Shoving, D16. Manhole Drop-off, D17. Patching Deterioration, D18. Raveling. Figure 1 illustrates the graphical display of D1 (Alligator Cracking) value distribution in the form of bar chart. The first bar is for instance: [0, 0.9) (644) represents the values of 644 records (of 2,115 records) are between 0 to 0.99. Figure 2 shows the statistics for D2 (Block Cracking). The minimum value is 0 and the maximum value is 8.

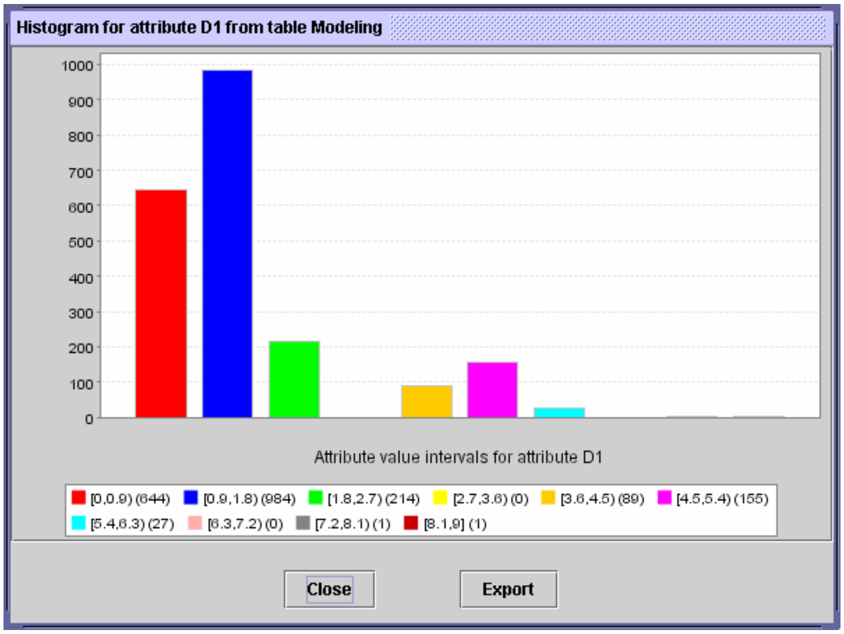

Figure 1: Attribute Distribution for D1 (Alligator Cracking)

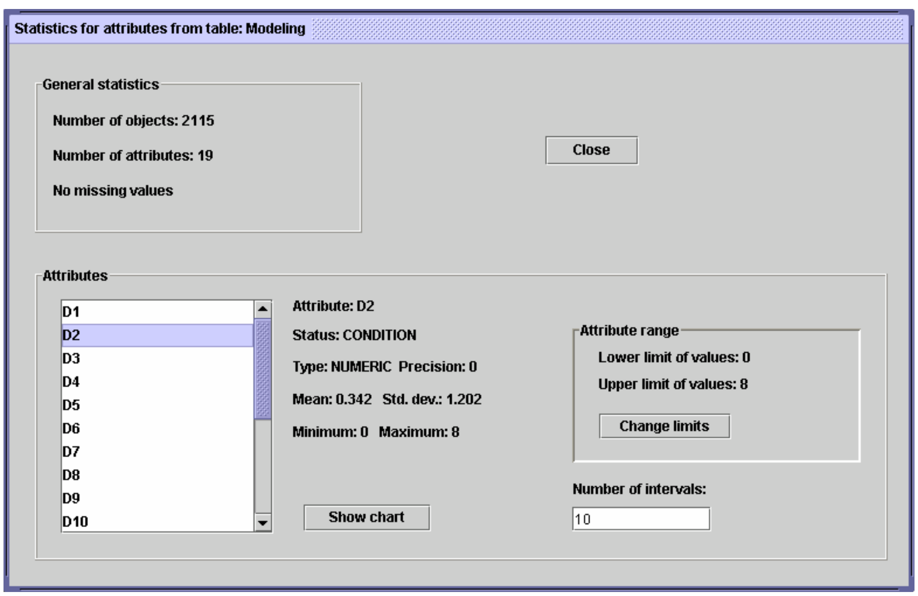

Figure 2: Information on Data Table (D2 for instance)

- The severity levels of distress are classified as L (low), M (moderate), and H (high). The coverage levels of distress are classified as A (local), B (medium), and C (extensive). Therefore, there are nine combinations (LA, LB, LC, MA, MB, MC, HA, HB, HC) of severity and coverage which are represented by number 1 to 9 respectively, and plus 0 represents no distress. For example, “D1 = 2” denotes Alligator Cracking occurts with low severity and medium coverage.

- M&R strategies used in the empircial study are classified as four types, which are represented as number 1 to 4 referring to no M&R required, preventive M&R, localized M&R and global M&R respectively.

M&R RULES INDUCTION AND VERIFICATION

This section will illustrate the RST application on M&R strategy induction and rules verification. Rough Set Exploration System (RSES) is employed to execute analyses. RSES is a software tool that provides the means for analysis of tabular data sets with use of various methods, in particular those based on RST (Warsaw University 2005).

Rules induction

Firstly, M&R rules are induced by using 2,115 records before attributes are deducted. The induced 435 rules are shown in Table 2. The first and second row in Table 2 represent the first and second rule, which match with 419 records (most records) and 205 records, respectively:

Table 2: The Induced 435 Rules with No Attribute Deduction

Rule set: Modeling

(1-435)	Match	Decision rules
1	419	(D1=1)&(D2=0)&(D3=0)&(D4=0)&(D5=0)&(D6=0)&(D7=0)&(D8=0)&(D10=0)&(D13=0)&(D14=0)=>(M&R={1[408],2[11]})
2	205	(D1=1)&(D5=1)=>(M&R={1[205]})
3	176	(D1=1)&(D2=0)&(D3=1)&(D4=0)&(D5=0)&(D6=0)&(D7=0)&(D8=0)&(D10=0)&(D13=0)=>(M&R={1[174],2[2]})
4	128	(D1=0)&(D2=0)&(D3=0)&(D4=1)&(D5=0)=>(M&R={2[128]})
5	83	(D1=2)&(D3=1)=>(M&R={1[83]})
6	67	(D1=1)&(D10=1)=>(M&R={1[67]})
7	62	(D1=0)&(D2=0)&(D3=1)&(D4=0)&(D5=0)&(D6=0)&(D8=0)&(D10=0)=>(M&R={1[1],2[61]})
8	55	(D1=5)&(D2=0)&(D3=0)&(D4=0)&(D5=0)&(D7=0)&(D8=0)&(D10=0)&(D13=0)=>(M&R={1[2],2[34],3[19]})
9	50	(D1=2)&(D2=0)&(D3=0)&(D4=0)&(D5=0)&(D6=0)&(D10=0)&(D11=0)&(D13=0)&(D16=0)=>(M&R={1[49],2[1]})
10	50	(D1=2)&(D2=0)&(D3=0)&(D4=0)&(D5=0)&(D8=0)&(D10=0)&(D11=0)&(D13=0)&(D16=0)=>(M&R={1[49],2[1]})
11	46	(D1=2)&(D5=1)=>(M&R={1[46]})
12	38	(D1=0)&(D2=0)&(D3=0)&(D4=0)&(D5=1)&(D6=0)&(D8=0)&(D10=0)&(D14=0)=>(M&R={1[1],2[21],3[16]})
13	33	(D1=0)&(D2=0)&(D3=1)&(D4=1)&(D5=0)&(D6=0)&(D13=0)=>(M&R={2[29],3[4]})
14	32	(D1=4)&(D2=0)&(D3=0)&(D4=0)&(D5=0)&(D7=0)&(D8=0)&(D13=0)=>(M&R={2[23],3[9]})
15	28	(D1=0)&(D3=1)&(D5=1)=>(M&R={2[28]})
16	28	(D1=0)&(D3=1)&(D4=1)&(D5=0)&(D6=0)&(D7=0)&(D13=0)=>(M&R={2[27],3[1]})
17	26	(D1=1)&(D2=0)&(D3=0)&(D4=1)&(D5=0)&(D7=0)=>(M&R={1[23],2[3]})
18	25	(D1=0)&(D3=5)&(D4=1)&(D7=5)=>(M&R={2[2],3[23]})
19	25	(D1=0)&(D3=5)&(D7=5)&(D8=0)=>(M&R={2[2],3[23]})
20	25	(D1=0)&(D3=5)&(D7=5)&(D10=0)=>(M&R={2[2],3[23]})
21	24	(D1=1)&(D8=1)=>(M&R={1[24]})
22	22	(D1=1)&(D2=1)=>(M&R={1[22]})
23	21	(D1=0)&(D3=1)&(D6=0)&(D10=1)=>(M&R={2[21]})
24	20	(D1=0)&(D2=5)&(D3=5)&(D7=5)=>(M&R={2[1],3[19]})
25	20	(D1=0)&(D2=5)&(D4=1)&(D7=5)=>(M&R={2[1],3[19]})
26	20	(D1=0)&(D2=5)&(D7=5)&(D10=0)=>(M&R={2[1],3[19]})
27	17	(D1=5)&(D2=5)&(D5=0)&(D6=0)=>(M&R={2[17]})
28	17	(D1=0)&(D2=0)&(D3=5)&(D4=0)&(D5=0)=>(M&R={2[17]})
29	17	(D1=0)&(D2=0)&(D3=5)&(D4=0)&(D7=0)=>(M&R={2[17]})
30	16	(D1=2)&(D10=1)=>(M&R={1[16]})
31	16	(D1=0)&(D2=1)&(D3=0)=>(M&R={2[16]})
32	15	(D1=2)&(D3=2)&(D7=0)=>(M&R={1[15]})
33	15	(D4=0)&(D5=6)&(D6=0)=>(M&R={2[15]})
34	15	(D1=0)&(D3=0)&(D4=1)&(D5=1)=>(M&R={2[8],3[7]})
35	13	(D1=1)&(D7=1)=>(M&R={1[13]})
36	13	(D1=0)&(D3=0)&(D4=4)&(D8=0)=>(M&R={2[13]})
37	13	(D1=0)&(D2=0)&(D3=0)&(D5=0)&(D10=1)=>(M&R={2[12],3[1]})
38	13	(D1=0)&(D2=0)&(D3=5)&(D4=1)=>(M&R={2[8],3[5]})
39	12	(D1=5)&(D2=5)&(D3=5)&(D6=0)=>(M&R={2[12]})
40	12	(D1=0)&(D5=1)&(D10=1)=>(M&R={2[12]})
41	12	(D1=5)&(D2=5)&(D4=5)=>(M&R={2[12]})
42	12	(D1=0)&(D3=0)&(D4=2)&(D5=0)&(D11=0)=>(M&R={2[11],3[1]})
43	12	(D1=0)&(D3=0)&(D4=2)&(D11=0)&(D14=0)=>(M&R={2[11],3[1]})
44	12	(D1=0)&(D2=1)&(D3=1)&(D4=0)&(D8=0)&(D10=0)=>(M&R={2[11],3[1]})

- If (D1=1) & (D2=0) & (D3=0) & (D4=0) & (D5=0) & (D6=0) &(D7=0) & (D8=0) & (D9=0) & (D10=0) & (D11=0) & (D12=0) & (D13=0) & (D14=0)

(That is, Alligator Cracking occurs with low severity and local coverage, and no other distress occurs.)
Then the adopted M&R strategy might be no M&R required (448 records) or preventive M&R (11 records)

- If (D1=1) & (D5=1)
  (That is, both Alligator Cracking occurs with low severity and local coverage and Edge Cracking occurs with low severity and local coverage.)
  Then the adopted M&R strategy will be no M&R required (205 records)

Then, RST is utilized to carry out reduct analysis for 18 attributes. Two groups of reduct set with 14 size attributes are shown below:

- D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D16
- D1, D2, D3, D4, D5, D6, D7, D8, D10, D11, D13, D14, D15, D16

The 13 attributes (cores) obtained from intersection of the above two reduct sets are D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D16. It is found that D12. Depression, D17. Patching Deterioration and D18. Raveling are not shown in both reduct sets.

M&R rules are induced again by using 2,115 records based on the reducted attributes. We could obtain 332 rules. Fewer rules are obtained than what of before attributes reduction. It demonstrated that RST can eliminate redundant attributes to improve the efficiency of rule induction. The rule which matches with the most records (208 records) is shown below:

If (D1=1) & (D2=0) & (D3=0) & (D4=0) & (D5=0) & (D6=0) & (D7=0) & (D8=0) & (D9=0) & (D10=0) & (D11=0) & (D12=0) & (D13=0) & (D16=0)
(That is, Alligator Cracking occurs with low severity and local coverage.)
Then adopted M&R strategy might be no M&R required (204 records) or fragment M&R required (4 records)

In addition, RST is able to carry out the shortening of reducts. One shortening ratio has to be assigned. This coefficient is between 0 and 1, which determines how "aggressive" the shortening procedure should be. The coefficient equals to 1.0 means that no shortening occurs. If shortening ratio is near zero, RST attempts to maximally shorten reducts. In the study, four shortening ratios - 1.0 (no shortening), 0.9, 0.7, and 0.5 - are selected to establish four shorten models for further verification and then 332, 89, 72, and 58 rules are obtained, respectively.

Rules verification

Additional 233 records are used to conduct verification analyses in the light of three shorten models and one no-shorten model. The results are shown in Table 3 to 6 and illustrated by Table 4:

- Rows in the Table correspond to actual decision classes (four M&R strategies) while columns represent decision values as returned by classifier in discourse. In Table 4, we may see that the constructed classifier sometimes mistakes objects from class 2 (preventive M&R) for those from class 3 (localized M&R) (2 such cases). We may also see that the classifier mistakes class 3 (localized M&R) for class 2 (preventive

M&R) few times more frequently than the other way round (6 cases as compared to 2).

- The values on diagonal represent correctly classified cases. If all non-zero values in the Table appear on the diagonal, we conclude that classifier makes no mistakes for a given set of data.
- No. of obj.: Number of objects in data set that belong to the decision class (M&R strategies) corresponding to current row. For instance, 135 represents there are 135 records out of 233 verification records belong to M&R strategy 1 (no M&R required).
- Accuracy: Ratio of correctly classified objects from the class to the number of all objects assigned to the class by the classifier. For instance, 0.898=53/(4+53+2), 53 represents the number of records whose actual M&R strategy is 2 (preventive) as the same with the predicted M&R strategy by 178 rules. If the pridicted strategy is 1 (4 records) and 3 (2 records), this must be incorrect.
- Coverage: Ratio of classified (recognized by classifier) objects from the class to the number of all objects in the class. That is, for all M&R strategies, the ratio of rules which can be recognized to carry out prediction (including incorrect prediction). For instance, 0.993=134/153; 0.797=(4+53+2)/74.
- True positive rate: For each M&R strategies, the ratio of strategy which can be correctly predicted by 178 rules. For instance, 0.97=134/(134+4); 0.88=53/(53+6+1).
- Total number of tested objected: Number of test objects used to obtain this result.
- Total accuracy: Ratio of number of correctly classified cases (sum of values on diagonal in Table) by 178 rules to the number of all tested cases (as in previous point). For instance, 0.936 = (134+53+2+2)/(134+4+53+6+1+2+2+2).
- Total coverage: Total coverage equals 1 which means that all objects have been recognized (classified) by 178 rules. Such total coverage is not always the case, as the constructed classifier may not be able to recognize previously unseen object. If some test objects remain unclassified, the total coverage value is less than 1. For instance, 0.876 = (134+4+53+6+1+2+2+2)/233.

By observing Table 3 to 6, it is found that shortening ratios increase (1.0 to 0.5) with rules decrease (332 to 58). The total accuracy is optimal (0.936) at 178 rules (Table 4). It is noted that 89 rules (shortening ratios = 0.9) has higher total accuracy than 332 rules (shortening ratios = 1.0). In addition, total coverage is optimal in Table 5 and 6. That represents that more generalized rules will be inducted and be more suitable for all new records after attribute reduction and reduct shortening. Hence the total coverage can then be improved. However, total accuracy must be defined depending on the correct M&R strategy prediction.

CONCLUSION AND SUGGESTION

Pavement distress survey can assist engineers in M&R strategy decisions. Proper M&R strategy can save long-term expense and keep the pavement above an acceptable serviceability. RST is used to induce M&R strategies by using 2,348 actual data (2,115 records for rule induction, and 233 records for rule verification). On the basis of the verification results, total accuracy for the induced rules is as high as 93.6%, which illustrates that RST certainly can assist engineers to easily remove redundant records, reduce attributes, discover association, and induce the most appropriate M&R strategy when they deal with the enormous pavement distress data. In the future, the database can be further extended for inducing more proper M&R strategies.

Table 3: Verification Results of Shortening Ratio = 1.0 (No Shortening, 332 Rules)

		Predicted						
		1	2	3	4	No. of obj.	Accuracy	Coverage
Actual	1	123	0	1	0	135	0.992	0.919
	2	3	45	15	0	74	0.714	0.851
	3	0	0	15	0	18	1	0.833
	4	0	0	0	5	6	1	0.833
	True positive rate	0.98	1	0.48	1			
Total number of tested objects: 233 Total accuracy: 0.908 Total coverage: 0.888								

Table 4: Verification Results of Shortening Ratio = 0.9 (89 rules)

		Predicted						
		1	2	3	4	No. of obj.	Accuracy	Coverage
Actual	1	134	0	0	0	135	1	0.993
	2	4	53	2	0	74	0.898	0.797
	3	0	6	2	0	18	0.25	0.444
	4	0	1	0	2	6	0.667	0.5
	True positive rate	0.97	0.88	0.5	1			
Total number of tested objects: 233 Total accuracy: 0.936 Total coverage: 0.876								

Table 5: Verification Results of Shortening Ratio = 0.7 (72 Rules)

		Predicted						
		1	2	3	4	No. of obj.	Accuracy	Coverage
Actual	1	133	1	0	1	135	0.985	1
	2	3	69	2	0	74	0.932	1
	3	0	16	2	0	18	0.111	1
	4	0	5	0	1	6	0.167	1
	True positive rate	0.98	0.76	0.5	0.5			
Total number of tested objects: 233 Total accuracy: 0.88 Total coverage: 1								

Table 6: Verification Results of Shortening Ratio = 0.5 (58 rules)

	Predicted							
Actual		1	2	3	4	No. of obj.	Accuracy	Coverage
	1	135	0	0	0	135	1	1
	2	47	27	0	0	74	0.365	1
	3	9	9	0	0	18	0	1
	4	5	0	0	1	6	0.167	1
	True positive rate	0.69	0.75	0	1			
Total number of tested objects: 233 Total accuracy: 0.7 Total coverage: 1								

ACKNOWLEDGMENTS

This study is partial results of project NSC 93-2211-E-159-003. The authors would like to express their appreciations to National Science Council of Taiwan for funding support.

REFERENCES

Attoh-Okine, N.O. (1997). “Rough Set Application to Data Mining Principles in Pavement Management Database.” *Journal of Computing in Civil Engineering*, 11 (4) 231-237.

Attoh-Okine, N.O. (2002). “Combining Use of Rough Set and Artificial Neural Networks in Doweled-pavement-performance Modeling - A Hybrid Approach.” *Journal of Transportation Engineering*, ASCE, 128 (3) 270-275.

Dubois, D. and Prade, H. (1991). In Z. Pawlark (Ed.), *Rough sets: Theoretical Aspects of Reasoning about Data*. Dordrecht, The Netherlands: Kluwer.

Grzymala-Busse, J.W. (1988). “Knowledge Acquisition under Uncertainty - A Rough Set Approach.” *Journal of intelligent and Robotic Systems*, 1 (1) 3-16.

Huang, Y.H. (1993). *Pavement Analysis and Design*. Prentice Hall Inc.

Pawlak, Z. (1982). “Rough Set.” *International Journal of Computer and Information Science*, 11 (5) 341-356.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R., and Ziarko, W. (1995). “Rough Sets.” *Communications of the ACM*, 38 (11) 88-95.

Radzikowska, A.M. and Kerre, E.E. (2002). “A Comparative Study of Fuzzy Rough Sets.” *Fuzzy Sets and Systems*, 126 (2) 137-155.

Walczak, B. and Massart, D.L. (1999). “Rough Sets Theory.” *Chemometrics and Intelligent Laboratory Systems*, 47 (1) 1-16.

Warsaw University (2005). Available at http://logic.mimuw.edu.pl/~rses.