The 25th International Symposium
on Automation and Robotics in Construction

June 26–29, 2008

Institute of Internet and Intelligent Technologies
Vilnius Gediminas Technical University
Saulėtekio al. 11, 10223 Vilnius, Lithuania
http://www.isarc2008.vgtu.lt/

ISARC-2008

USING DATA MINING TO EXPLORE THE DETERIORATION FACTORS OF REINFORCED CONCRETE (RC) HIGHWAY BRIDGES IN TAIWAN

Rong-Yau Huang
Professor, Graduate Inst. of Constr. Eng. & Mgmt, National Central U.,Taoyuan, Taiwan
rhuang@cc.ncu.edu.tw

I-Shiang Mao
PhD student, Graduate Inst. of Constr. Eng. & Mgmt, National Central U.,Taoyuan, Taiwan
963405002@cc.ncu.edu.tw

Hao-Kang Lee
Master, Graduate Inst. of Constr. Eng. & Mgmt, National Central U., Taoyuan, Taiwan
concon@ms34.url.com.tw

ABSTRACT

Information about the factors that lead to the deterioration of bridges is essential for bridge maintenance. Pinpointing what these factors are will certainly enhance the effectiveness of bridge management. However, a review of the literature reveals that such deterioration factors are usually determined from expert opinion. In other words, there is no systematic way to identify the factors and the effect they have on different types of bridge members. This study identifies six common types of deterioration that affect RC bridge decks. Twenty-nine factors are extracted from a review of past related work as well as from the inventory of Taiwan Bridge Management System. After this, a data mining technique, Rough Set Theory (RST), is employed to find the factors that have the greatest impact on deterioration from thousands of visual inspection, traffic and environmental data. It is found that weather-related factors are rather significant for almost all types of deterioration. In addition to these, some functional and structural factors are major factors for cracking and traffic volume is a major factor in rebar corrosion and breakage.

KEYWORDS

Bridges, deterioration, factor, data mining, rough set

1. INTRODUCTION

According to the statistics collected by the Taiwan Bridge Management System (TBMS), nearly 50% of the bridges in this country are aged 20 years or more, making bridge management and maintenance become increasingly important. The factors that have an impact on bridge deterioration provide essential information for bridge management and maintenance. Pin-pointing these factors will most certainly help to enhance the effectiveness of bridge management as well as to estimate their life cycle cost. For instance, bridges could be categorized for managerial purposes according to their common deterioration factors. Information about the degree of deterioration of bridges in the same category could be used by the bridge administration to prioritize the sequence of maintenance. Given budget limitations, a fair and reasonable maintenance policy could then be proposed. The identification of deterioration factors could also help in the modelling of bridge deterioration, which in turn improves forecasting.

Deterioration factors can include many sources such as weather, traffic volume, nature of the design, quality of the construction, and so on. Most investigative efforts so far have either focused on experimenting with some preselected factors, or factors suggested by experts. There have been a few studies where statistical methods have been applied to determine deterioration factors from preselected ones but these however, have targeted only a few factors. Worries about missing important factors still remain. Reviews of past literature reveal a lack of a systematic way to identify factors for different types of bridge member deterioration.

The technique of data mining has already been widely employed for factor identification in a variety of other fields. One of the advantages of data mining over the traditional statistical methods is the ability to find implicit information or characteristics from large amounts of accumulated data through systematic mathematical algorithms. In this study we apply Rough Set Theory (RST), one of the most common of the data mining techniques, to identify the factors that lead to deterioration of bridges. We first focus on some common types of RC bridge deck deterioration, such as cracking, spalling, efflorescence, etc. We select 29 factors from a review of past related work as well as the TBMS inventory. The relevant factors are then explored by applying RST to each deterioration group.

2. LITERATURE REVIEW

2.1. Research on Bridge Deterioration

The deterioration factors for the whole bridge are usually taken into account. Scherer [1] established a Markov-chain decision model to diagnose the overall situation for bridge deterioration, in which road level, average annual rainfall, traffic volume, bridge material, number of spans, and bridge age were chosen as factors. Zhao [2] proposed a fuzzy system for concrete bridge damage diagnosis while considering structural type, span length, number of lanes, number of spans, paving material, average annual rainfall, temperature variation, traffic volume and road level. Su [3] utilized logistic regression analysis to show how bridges deteriorated due to functional, structural and environmental factors. Chang [4] developed a model of bridge deterioration after consulting the relevant literature and summarizing the factors that could cause deterioration such as, age, structural type, traffic volume, local soil profile, average annual rainfall, road level, and seismic zone. Huang [5] described the deterioration trend after screening bridge data, matching similar environmental conditions such as whether the bridge was over water or not, distance from coastlines, average annual rainfall, soil profile and seismic zone, etc.

There have been few studies focused on bridge components. Huang [6] did use an Analysis of Variation (ANOVA) technique to determine the factors that could have an influence on bridge deterioration, such as the number of spans, area of the bridge deck, location, length of the bridge, average daily traffic volume, designed loading, and whether over water or not. Chen [7] proposed a model of the deterioration of the eight most commonly damaged bridge components. The major factors leading to deterioration of these bridge components were obtained from the related literature as well as consultation with bridge experts. A summary of the factors from past studies can be found in Table 1. These efforts lead us to the following conclusions:

Most studies have targeted factors affecting the whole bridge rather than its component parts.

Consulting the literature and expert opinion have been the major sources when seeking factors.

Different types of deterioration have not yet been taken into account.

Factors discussed vary from study to study, so work covering most factors is very rare.

Undoubtedly, it would be more sensible to study how factors affect bridge components rather than treating the bridge as a whole, and grouping the deterioration types will give us a more accurate understanding of bridge deterioration. In addition, some factors can be less than obvious, so may be overlooked if a systematic examination is not carried out. In this study, we demonstrate an effective and efficient approach to explore the factors leading to the deterioration of RC bridge decks.

2.2. Application of Data Mining

Data mining is a powerful new technology with great potential to extract hidden predictive information from large databases [8]. Data mining tools have already been widely applied to pattern recognition or data classification as well as to help with predicting future trends or behaviours.

Table 1. Summary of Studies on Bridge Deterioration [1–7]

Targets		Whole Bridges					Components	
Factors \ Studies		Scherer (1994)	Zhao (2002)	Su (2003)	Chang (2004)	Huang (2005)	Huang (2003)	Chen (2005)
Functional	Bridge Age	✓		✓	✓			✓
	No. of Spans	✓	✓				✓	
	No. of Lanes		✓	✓				✓
	Length of Bridge		✓	✓			✓	✓
	Area of Deck						✓	✓
	Max. Span		✓	✓				
Structural	Structural Type		✓	✓	✓			✓
	Girder Type			✓				✓
	Girder Material	✓						
	Abutment			✓				
	Pavement		✓	✓				
	Earthquake Bracing			✓				
	Expansion Joint							✓
	Wing wall			✓				
	Designed Live Load						✓	✓
Traffic	Traffic Volume	✓	✓		✓		✓	✓
Environmental	Over Water or Not?			✓		✓	✓	
	Distance from Coast			✓		✓		✓
	Acid Rain							✓
	Avg. Yearly Rainfall	✓	✓		✓	✓		
	Avg. Rainy Days per Year							✓
	Soil Profile				✓	✓	✓	
	Temperature Variation		✓					
Misc.	Road Level	✓	✓		✓			
	Seismic Zone				✓	✓		

With today's highly efficient mathematical algorithms, and computer software, we can scour large databases for hidden patterns, finding predictive information that lies outside the experts' expectations. That is, data mining can find answers to questions, which traditional statistical tools will take too long to reach, especially when working from large databases. Many computer software tools have been developed for solving such problems. We now mention some successful applications: Wu [9] applied K-means clustering and a decision tree to identify the characteristics of drivers with high accident rates, in an effort to help insurers screen customers; Lin [10] utilized RST to determine the factors leading to the deterioration of highway pavement; Liu [11] built a disease classification model for abdominal diseases by using RST to take out redundant attributes. One can see that this new technology matured enough to be widely applied in a variety of fields.

3. METHODOLOGY

3.1. K-Means Clustering

K-means Clustering, first proposed by MacQueen [12], is one of the simplest unsupervised learning algorithms. The main idea is to classify a given data set into a certain number of clusters. This is useful for pre-processing when the volume of data is large and discrete. First, k centroids are defined, one for each cluster, preferably placed as far away from each other as possible. Next, each data point is assigned to a given data set associated with the nearest centroid. To do this, the Euclidean distance $d(Xi,Cj)$ between each data point and the centroid is first calculated

$$d(Xi, Cj) = \left(\sum_{d=1}^{n} \left| Xi_d - Cj_d \right|^2 \right)^{1/2}, \tag{1}$$

where n is the dimension of the data.

The first run is completed and an initial clustering is obtained when there is no data point left pending assignment. Next, new centroids are recalculated based on the clusters resulting from the previous step. New clustering is done by repeating the steps mentioned above. The process continues until the centroids no longer move. Finally, we use this algorithm to minimize the objective function:

$$E = \sum_{i=1}^{k} \sum_{x \in S_i} \left\| x - c_i \right\|^2, \tag{2}$$

where S_i and c_i are the data set and the centroid of cluster i, respectively. Generally speaking, the algorithm includes the following steps:

Assign K points to the space represented by the data set being clustered.

Assign each data point to the group with the closest centroid.

Recalculate the positions of the K centroids.

Repeat Steps 2 and 3 until the centroids no longer move.

The procedure is shown in Figure 1

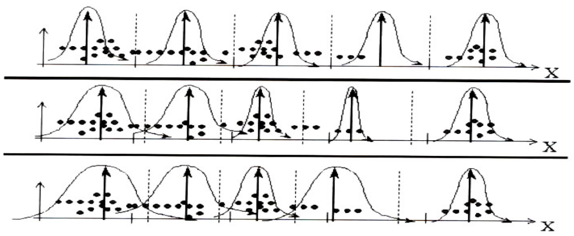


Figure 1. Procedure for K-Means Clustering

3.2. Rough Set Theory

Rough Set Theory, one of most popular data mining tools, was first proposed by Pawlak [8] as a new mathematical tool for dealing with vague information. The basic concept on which RST functions is the notion of an approximation space, which is an ordered pair $S=(U,R)$, where U is a non-empty and finite set of objects, called a universe; R is the equivalence relation on U, called the indiscernibility relation. Each equivalence class induced by R is called an elementary set in S. A definable set in S is any finite union of elementary sets in S. For each X $\subseteq$ U, X can be characterized in S on relation R by a pair of sets, its lower and upper approximations in S, defined as

$$R_{\text{low}}(X) = \{x \in U \mid [x]_R \subseteq X\}, \tag{3}$$

$$R_{\text{upp}}(X) = \{x \in U \mid [x]_R \cap X \neq \varnothing\}, \tag{4}$$

where $[x]_R$ denotes the equivalence class of R containing x. A rough set in S is the family of all subsets of U having the same lower and upper approximations.

Likewise, an information system (attribute-value system) is a pair $I=(U,A)$ where U is a non-empty, finite set of objects and A is a non-empty, finite set of attributes on U, such that $fa{:}U \rightarrow Va$ for every attribute $a \in A$. Va is the set of values that a may take. In the Rough Set framework data are represented in the form of an information table. Each row of the table represents an object and every column represents an attribute that can be measured for each object. In other words, the information table simply assigns a value in Va to each attribute a of each object in U. With any B $\subseteq$ A, there will be sets of objects that are indiscernible based on those attributes. These indistinguishable sets of objects define an equivalence relation, called the B-Indiscernibility relation, defined as follows:

$$\mathrm{IND}(B)=\{(x,x')\in U^2 \mid \forall a\in B, f_a(x)=f_a(x')\}. \quad (5)$$

Similarly, the target set $B(X)$ can be approximated using only the information contained within B by pegging the lower and upper approximations of set X

$$B_{\mathrm{low}}(X)=\{x\in U \mid [x]_{IND(B)}\subseteq X\}, \quad (6)$$

$$B_{\mathrm{upp}}(X)=\{x\in U \mid [x]_{IND(B)}\cap X\neq\varnothing\}. \quad (7)$$

The resultant boundary region is given by the set difference $B_{\mathrm{upp}}(X)$ - $B_{\mathrm{low}}(X)$, which consists of those objects that can neither be ruled in nor ruled out as members of the target set X. The concept of lower and upper approximations is illustrated in Figure 2.

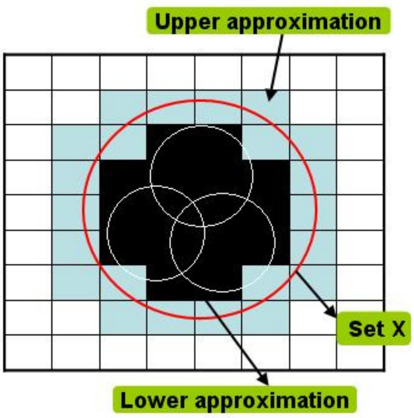


Figure 2. Approximations of Set X

In general, the upper and lower approximations are not equal. In such cases, set X is undefinable or roughly definable on attribute set B. When the upper and lower approximations are equal (i.e., the boundary is empty), then set X is definable on attribute set B.

If there is an attribute set $C \subseteq B$, which by itself can fully characterize the information system based on attribute set B, such an attribute set C is called a reduct. Formally, a reduct C is a subset of attributes B such that [13]: $[x]C = [x]B$;

Attribute set C is minimal in the sense that $[x](C\text{-}a) \neq [x]B$ for any attribute $a \in C$.

In other words, no attribute can be removed from a reduct without changing the equivalence classes $[x]_B$. The reduct of an information system may not be unique. There may be many subsets of attributes B which hold the equivalence class structure implied in the information system. The set of attributes which is common to all reducts is called the core. Finding reducts and their cores is very useful. The elimination of attributes will minimize the consumption of computer resources without changing the nature or characteristics of the information system. This is what makes RST so efficient for dealing with a large volume database.

To verify the result of analysis, 10% of the data are normally preserved as testing data. The accuracy of the analysis is then calculated by (A+D)/(A+B+C+D), where A, B, C and D denote:

Actual \ Predicted	1	0
1	A	B
0	C	D

3.3. Process of Analysis

The major steps in the analysis used in this study to find corresponding factors which cause RC bridge deck deterioration, are summarized below:

All the factors under consideration are listed. Factors obtained from a review of past studies and the attributes associated with bridge features in the TBMS are examined as extensively as possible.

Data are then pre-processed for data mining. Categorical data (e.g., type of pavement) are simply converted to ordinal numbers, while numerical data (e.g., average annual rainfall) are categorized using K-Means Clustering.

Data mining is conducted. The relevant factors leading to the deterioration of bridge decks are explored by RST.

The results are validated by comparing them with independent statistical tests. It is proposed that the Mann-Whitney U (MWU) test be utilized to assess the correlation between identified factors and their corresponding deterioration type.

A flowchart of the analysis process is shown in Figure 3.

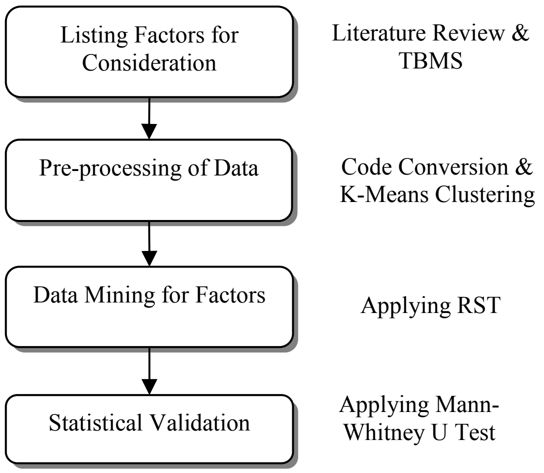


Figure 3. Flowchart of the Analysis

To facilitate the analysis, computer programs such as RSES (Rough Set Exploration System) and SPSS (Statistical Package for the Social Sciences) are used for data mining, K-Means Clustering and statistical testing.

4. DATA TREATMENT

4.1. Data Collection

The factors leading to the deterioration of bridge decks can be divided into four categories: functional, structural, environmental, and traffic. We identify 29 factors from past studies and the attributes listed in the TBMS. These factors are summarized in Table 2. We gather data from a total of 2,128 bridges in the Taiwan National Freeway System. Both the values of the attributes and the visual inspection records for each bridge are collected.

The analysis now focuses on the six most commonly seen types of deterioration: cracking, spalling, efflorescence, corrosion of rebar, and breakage and honeycombing of the bridge decks. Samples for analysis are sorted into the six types. In TBMS, bridge conditions are assessed in terms of a 0 to 4 rating scale for the degree, extent, and relevancy (DER) by inspectors (see Table 3). All samples with a D≥2 rating are considered as deteriorated in this study. The final samples of each type are summarized in Table 4.

Table 2. Candidate Deterioration Factors

No.		Functional Attributes
1	▲	Bridge Age
2	▲	No. of Spans
3	▲	No. of Lanes
4	▲	Length of Bridge
5	▲	Area of Bridge Deck
6	▲	Length of Max. Span
		Structural Attributes
7	■	Structural Type
8	■	Type of Pavement
9	■	Type of Girder
10	■	Type of Girder Material
11	■	Type of Expansion Joint
12	■	Type of Bracing for Earthquake
13	■	Type of Bearing
14	■	Designed Live Load
15	■	Type of Pier
16	■	Type of Pier Material
17	■	Type of Pier Foundation
18	■	Type of Abutment
19	■	Type of Abutment Foundation
20	■	Type of Wing Wall
		Environmental Attributes
21	■	Whether Over Water or Not?
22	▲	Distance from Coast
23	■	Significance of Acid Rain
24	▲	Avg. Annual Rainfall
25	▲	Peak Monthly Rainfall
26	▲	Avg. Rainy Days per Year
27	▲	Max. Rainy Days in a Month
28	■	Soil Profile
		Traffic Attributes
29	▲	Avg. Annual Traffic Volume

■Categorical Data ▲Numerical Data

Table 3. DER Rating after Visual Inspection

	0	1	2	3	4
D	No such item	Good	Fair	Poor	Sever
E	Cannot be inspected	< 10% < 30% < 60% <			
R	Cannot be decided	Minor	Small	Medium	High

Table 4. No. of Samples Showing Deterioration

Type of Deterioration	No. of Samples Showing Deterioration
Cracking	716
Spalling	99
Efflorescence	704
Corrosion of rebar	525
Breakage	858
Honeycombing	177

4.2. Code Conversion of Categorical Factors

As summarized in Table 2, 17 out of 29 factors are categorical; these factors are represented in a nominal scale. The samples are simply clustered into finite discrete categories. Obviously, further analysis is more convenient if each category is represented by a set of ordinal numbers. For instance, there are 6 girder types in the data inventory; each type is assigned a different number (see Table 5).

Table 5 Types of Girders and Code Conversion

Types of Girders	Code
I Section	1
T Section	2
U Section	3
Slab Beam	4
Box Section	5
Others	6

Code conversion can be done similarly for all other categorized factors. It should be noted that the ordinal numbers do not represent any interval or rating scale.

4.3. Clustering of Numerical Factors

Unlike categorical factors, numerical factors do have an interval or rating scale. The values of factors can cover a wide range when one is working with a huge amount of data. Therefore, a reasonable and objective way to partition the data set is required. We use a well-known algorithm, K-Means Clustering, to solve this problem. However, it is particularly troublesome to decide on the value of K, since the number of clusters is often arbitrary. There is no general theoretical solution to find the optimal number of clusters for a given data set. Our approach is to compare the results of multiple runs with different K and choose the best one according to a given criterion, i.e., the accuracy of the RST analysis. For example, the length of a maximum span is one of the numerical factors. Trial partitions of data set are shown in Table 6.

Table 6. Partitions for the Length of Max. Span

K	Centroids (m)	Partitions (m)
2	37	0~200
	346	>200
3	34	0~60
	90	60~300
	378	>300
4	34	0~60
	87	60~200
	293	200~400
	425	>400
5	27	0~35
	45	35~72
	100	72~200
	293	200~400
	425	>400

The code conversion for each partition with different K is carried out in a similar fashion. Afterwards, RST trial analysis for different K is conducted. The outputs for each case are shown in Fig.4-Fig.6. The trial process shows us that K=4 results in the best accuracy of analysis. Therefore, 4-partitions offer the best clustering for the length of maximum span factor. With this approach, we can find out which clustering solution best reflects the significant characteristics of the data. The other numerical factors are treated in the same way (Figures 4–6).

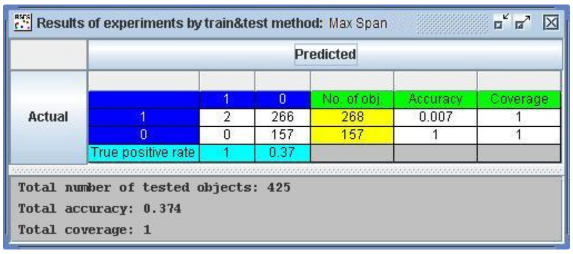


Figure 4. Accuracy for K=2 (0.374)

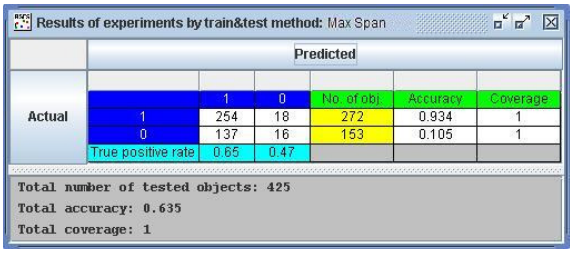


Figure 5. Accuracy for K=3 (0.635)

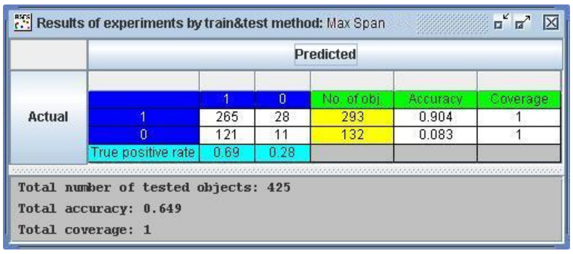


Figure 6A. Accuracy for K=4 (0.649)

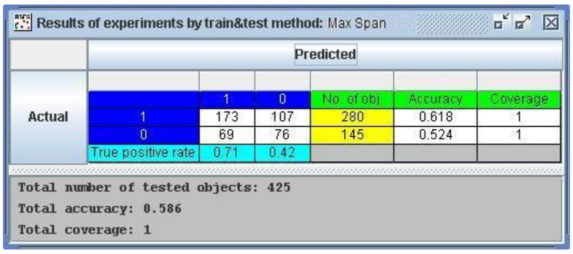


Figure 6B. Accuracy for K=5 (0.586)

5. EXPLORING FACTORS

5.1. Mining Factors by RST

After the raw data have been treated (i.e., clustering and code conversion), the next step is to extract the relationship between the factors and deterioration of the RC bridge decks. As mentioned above, the data are represented in the form of an information table. Each row of the table represents a bridge sample, and every column represents an attribute associated with that sample. Factors to be identified are defined as condition attributes, while the state of deterioration (i.e., whether D≥2 or not) is taken as the decision attribute. The information table is now called the decision table, and includes both the condition attributes and the decision attribute for each sample. Table 7 shows an example.

Table 7. Example of a Decision Table

U	Condition Attributes			Decision Attributes
	a_1	a_2	a_3	D≥2
x_1	2	1	3	0
x_2	3	2	1	1
x_3	2	1	3	1
x_4	3	2	2	0
x_5	1	1	4	0

In the table, x_1~x_5 indicate samples associated with 3 condition attributes a_1~a_3, for each attribute in the universe. It appears that all condition attributes between x_1 and x_3 are identical which however, results in a different decision attribute. In other words, there is a logically unacceptable inconsistency in the table. Any sample showing an inconsistency of this kind must be removed. This is done step-by-step until no conflict remains. The entries remaining from Table 7 are shown in Table 8.

Table 8. Decision Table without Inconsistencies

U	Condition Attributes			Decision Attributes
	a_1	a_2	a_3	D≥2
x_2	3	2	1	1
x_4	3	2	2	0
x_5	1	1	4	0

Eventually, a decision table without any inconsistency is obtained. Some attributes may not contribute to changes in the consistency of a table. For example, x_2, x_4 and x_5 are always consistent with each other regardless of whether attribute a_1 is neglected or not. That is, a_1 can be pruned without affecting the consistency of Table 8. In contrast, removing attribute a_3 does affect the consistency (i.e., resulting in an inconsistency between x_2 and x_4), and hence is not removable. The removal of a_1 and a_3 are shown in Table 9 and Table 10, respectively.

Table 9. Removal of a_1 in Table 8

U	Condition Attributes		Decision Attributes
	a_2	a_3	D≥2
x_2	2	1	1
x_4	2	2	0
x_5	1	4	0

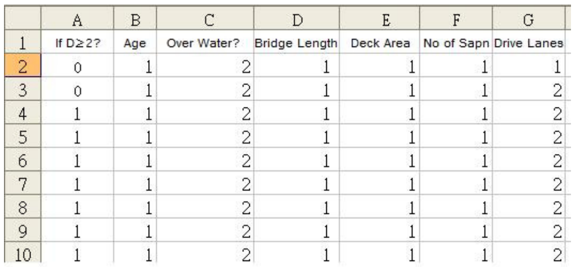
Table 10. Removal of a_3 in Table 8

U	Condition Attributes		Decision Attributes
	a_1	a_2	D≥2
x_2	3	2	1
x_4	3	2	0
x_5	1	1	0

A reduct is the set of remaining condition attributes after taking out the removable ones. Obviously, there are several paths whereby this can be done. The RST computer tools can go through all possible paths and produces all reducts. The set of attributes which is common to all reducts, called the core, represents the set of factors which cannot be ignored.

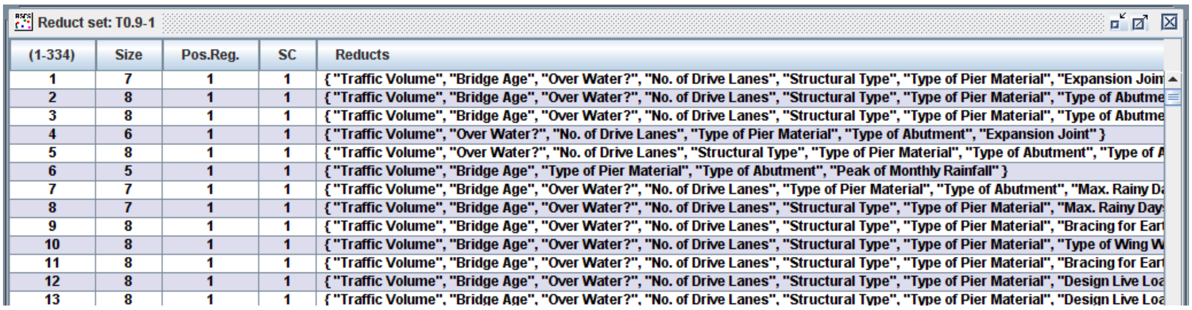
The degree of significance of factors can be measured by the reduct occurrence rate, where the higher the occurrence rate, the greater the significance. In this study, significant factors are those with an occurrence rate greater than 75% while the minor ones have a rate between 10-75%. The remainder, that is, occurrence rate less than 10%, are classified as "not related."

To facilitate the approach we utilize a well known RST tool developed by the University of Warsaw [14–15], RSES (Rough Set Exploration System v2.2.2), to conduct the analysis. We build the decision table shown in Figure 7. In column A, "if D≥2", indicates the decision attribute while the others are condition attributes. The reducts are then mined by RSES. An example of the output is shown in Figure 8. Six types of RC bridge deck deterioration are analyzed. Both major and minor factors are summarized in Tables 9-10.



	A	B	C	D	E	F	G
1	If D≥2?	Age	Over Water?	Bridge Length	Deck Area	No of Sapn	Drive Lanes
2	0	1	2	1	1	1	1
3	0	1	2	1	1	1	2
4	1	1	2	1	1	1	2
5	1	1	2	1	1	1	2
6	1	1	2	1	1	1	2
7	1	1	2	1	1	1	2
8	1	1	2	1	1	1	2
9	1	1	2	1	1	1	2
10	1	1	2	1	1	1	2

Figure 7. Decision Table



Reduct set: T0.9-1

(1-334)	Size	Pos.Reg.	SC	Reducts
1	7	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Expansion Join
2	8	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Type of Abutme
3	8	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Type of Abutme
4	6	1	1	{ "Traffic Volume", "Over Water?", "No. of Drive Lanes", "Type of Pier Material", "Type of Abutment", "Expansion Joint" }
5	8	1	1	{ "Traffic Volume", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Type of Abutment", "Type of A
6	5	1	1	{ "Traffic Volume", "Bridge Age", "Type of Pier Material", "Type of Abutment", "Peak of Monthly Rainfall" }
7	7	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Type of Pier Material", "Type of Abutment", "Max. Rainy Da
8	7	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Max. Rainy Day
9	8	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Bracing for Ear
10	8	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Type of Wing W
11	8	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Bracing for Ear
12	8	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Design Live Loa
13	8	1	1	{ "Traffic Volume", "Bridge Age", "Over Water?", "No. of Drive Lanes", "Structural Type", "Type of Pier Material", "Design Live Loa

Figure 8. Reducts Mined by RSES

Table 9. RC Bridge Deck Deterioration Factors

Deterioration Types	Major Factors Occurrence Rate ≥75%	Minor Factors 10% ≤ Occurrence Rate < 75%	
Cracking	● Peak Monthly Rainfall ● Max. Rainy days in a Month ● Type of Girder Material ● No. of Lanes ● Expansion Joint ● Type of Pier ● Whether Over Water or Not? ● Designed Live Load	● Bridge Age ● Length of Bridge ● Area of Bridge Deck ● No. of Spans ● Max. Span ● Structural Type ● Type of Pier Material ● Type of Pier Foundation ● Type of Girder ● Type of Abutment ● Type of Abutment Foundation ● Type of Wing Wall	● Type of Bearing ● Earthquake Bracing ● Soil Profile ● Distance from Coast ● Avg. Rainy Days per Year ● Avg. Annual Rainfall ● Traffic Volume
Efflorescence	● Peak Monthly Rainfall ● Max. Rainy days in a Month ● Type of Pier ● Soil Profile ● Type of Girder Material	● Bridge Age ● Whether over Water or Not? ● Length of Bridge ● Area of Bridge Deck ● No. of Spans ● No. of Lanes ● Max. Span ● Structural Type ● Type of Pier Material ● Type of Pier Foundation ● Type of Girder	● Type of Abutment ● Type of Abutment Foundation ● Expansion Joint ● Type of Wing Wall ● Type of Bearing ● Earthquake Bracing ● Designed Live Load ● Distance from Coast ● Avg. Rainy Days per Year ● Avg. Annual Rainfall ● Traffic Volume
Corrosion of Rebar	● Peak Monthly Rainfall ● Traffic Volume	● Bridge Age ● Whether over Water or Not? ● Length of Bridge ● Area of Bridge Deck ● No. of Spans ● No. of Lanes ● Structural Type ● Type of Pier ● Type of Pier Foundation ● Pavement Material ● Type of Girder ● Type of Girder Material	● Type of Abutment ● Type of Abutment Foundation ● Expansion Joint ● Type of Wing Wall ● Type of Bearing ● Earthquake Bracing ● Designed Live Load ● Soil Profile ● Distance from Coast ● Avg. Rainy Days per Year ● Avg. Annual Rainfall ● Max. Rainy days in a Month
Spalling	● Peak Monthly Rainfall	● Bridge Age ● Whether over Water or Not? ● Length of Bridge ● Area of Bridge Deck ● No. of Spans ● No. of Lanes ● Structural Type ● Type of Pier ● Type of Pier Foundation ● Type of Girder ● Type of Abutment ● Type of Abutment Foundation	● Expansion Joint ● Type of Wing Wall ● Type of Bearing ● Earthquake Bracing ● Designed Live Load ● Soil Profile ● Distance from Coast ● Avg. Rainy Days per Year ● Avg. Annual Rainfall ● Max. Rainy days in a Month ● Traffic Volume

Table 10. The Deterioration Factors of RC Bridge Decks (Continued)

Deterioration Types	Major Factors Occurrence Rate ≥75%	Minor Factors 10% ≤ Occurrence Rate < 75%	
Breakage	● Peak Monthly Rainfall ● Max. Rainy days in a Month ● Traffic Volume	● Bridge Age ● Whether over Water or Not? ● Length of Bridge ● Area of Bridge Deck ● No. of Spans ● No. of Lanes ● Max. Span ● Structural Type ● Type of Pier ● Type of Pier Foundation ● Type of Girder ● Type of Girder Material	● Type of Abutment ● Type of Abutment Foundation ● Pavement Material ● Expansion Joint ● Type of Wing Wall ● Type of Bearing ● Earthquake Bracing ● Designed Live Load ● Soil Profile ● Distance from Coast ● Avg. Rainy Days per Year ● Avg. Annual Rainfall
Honeycombing	● Peak Monthly Rainfall ● Max. Rainy days in a Month	● Bridge Age ● Whether over Water or Not? ● Length of Bridge ● Area of Bridge Deck ● No. of Spans ● No. of Lanes ● Max. Span ● Structural Type ● Type of Pier ● Type of Pier Material ● Type of Pier Foundation ● Type of Girder ● Type of Girder Material	● Type of Abutment ● Type of Abutment Foundation ● Expansion Joint ● Type of Wing Wall ● Type of Bearing ● Earthquake Bracing ● Designed Live Load ● Soil Profile ● Distance from Coast ● Avg. Rainy Days per Year ● Avg. Annual Rainfall ● Traffic Volume

5.2. Data Findings and Results

In common with most concrete structures, the six types of deterioration in this study can be caused by either physical or chemical attacks or even the interaction of both. Since the data collected from the historical visual inspection records in TBMS are based on inspections, these records reflect the viewpoints of the inspectors. For instance, "breakage" is the most common type of damage noted, but breakage could be the result of several types of events that would be seen and noted by the inspectors. However, "honeycombing" is usually caused by poor compaction or vibration during concrete casting and is repaired as soon as the form is stripped, so that it is rarely seen. Other types of deterioration usually follow a certain sequence. For example, "cracking" can be initially induced by traffic action but the opening of the crack can accelerate "efflorescence" and "corrosion of rebar" which are also affected by chemical reactions (i.e., leaching and carbonation). Finally, "spalling" appears when the rusting of the steel forces the concrete to crack. As the deterioration is in progress, the earlier degrading phenomenon (e.g., cracking) should be seen more than the later ones. While recalling the figures in Table 4, "breakage" and "cracking" account for the largest portion (i.e., 28% and 23% respectively) of total samples showing deterioration. However, "spalling" and "honeycombing" are only 6% and 3% respectively. Therefore, the distribution of samples for different types of deterioration in Table 4 can be considered as a result of an objective sampling.

Data mining indicates that environmental factors, mainly weather-related factors, have a significant effect on all types of deterioration. "Type of Pier," "type of girder material" and "traffic volume" are also rather significant. All major factors and their related deterioration types are summarized in Table 11.

Table 11 Major Factors for Each Deterioration Type

Factors \ Deterioration	Cracking	Efflorescence	Corrosion of Rebar	Spalling	Breakage	Honeycombing
Whether over Water or Not?	✓					
No. of Lanes	✓					
Type of Pier	✓	✓				
Type of Girder Material	✓	✓				
Type of Expansion Joint	✓					
Designed Live Load	✓					
Soil Profile		✓				
Peak Monthly Rainfall	✓	✓	✓	✓	✓	✓
Max. Rainy Days in a Month	✓	✓			✓	✓
Traffic Volume			✓		✓	

Obviously, more major factors in cracking or efflorescence than in other deterioration types are identified by RST. The outputs of RSES show that the reducts in cracking and efflorescence are larger than those in other deterioration types. The average size of reducts in cracking and efflorescence is 13 while it is about 8 in spalling and breakage. As a result, more major factors are produced as larger reducts have more factors in common with each other. It is noted that "peak monthly rainfall" and "Max. rainy days in a month" are more significant than "Avg. annual rainfall" or "Avg. rainy days per year." This suggests that intensive rainfall in a short period has a more significant affect on bridge decks deterioration.

In addition to weather-related factors, "whether over water or not?" "No. of lanes", "type of pier", "type of girder material", "expansion joints", and "designed live load" are identified as major factors related to "cracking". Comparison of samples showing deterioration or not provides that higher percentage of RC bridge decks are cracked if the bridges are over water, having 3 or more lanes, shored by multi-columns bent or the deck slabs are cast on steel girders. On the other hands, less bridge decks are found cracked while the traffic surcharge load is taken into account in design or modular type of expansion joint are used. This suggests that the action of traffic and selection of expansion joint has a great impact on cracking in bridge decks.

Some major factors in cracking, for example, "type of pier", "type of girder material", also turn up in efflorescence. Besides, the distributions of samples showing deterioration in both are found similar, too. This implies that efflorescence somewhat correlates with the occurrence of cracking in RC bridge decks. Unexpectedly, "soil profile" is recognized as a major factor contributing to efflorescence. In fact, more RC bridge decks are found to have this type of deterioration while the bridges were built on soft strata (e.g. Taipei basin).

"Traffic volume" is identified as a major factor leading to rebar corrosion and breakage. The number of samples showing deterioration suggests that there seems to be a positive correlation between traffic volume and the probabilities of deterioration in both types.

"Distance from coast" is usually regarded as a major factor affecting to rebar corrosion by intuition. However, it is not so significant in this study. This is understandable since all the bridge samples were collected from national freeways which are considerably away from coast. Chen [7] indicated that the salt damage is unobvious while the concrete structures are located more than 10 km away from coast.

5.3. Statistical Testing

The deterioration factors obtained through RST are compared with the results obtained from a statistical approach. The MWU, one of the best-known non-parametric significance tests, is employed to assess the correlation between the identified factors and corresponding deterioration type. Generally, the MWU is used for assessing whether two observations come from the same distribution. The null hypothesis is that the two samples are drawn from a single population, and therefore that their probability distributions are equal. In this study, we consider that a factor is significant, if the probability distributions between observations showing deterioration or non-deterioration are different. The null hypothesis, denoted as H_0, is stated as follows: "a factor is not related to a specified deterioration type". The level of statistical significance is 5%, which indicates the probability of the test statistic (i.e. U) falling into the rejecting region when the null hypothesis is correct. The MWU requires the two samples to be independent, and the observations to be either in ordinal or rating scale. In other words, the test is meaningful only when conducted on numerical factors. The MWU is performed by a commonly use computer software, SPSS. The MWU results and comparison with RST results are summarized in Table 12.

It appears that according to MWU most factors are related to deterioration. That is, the probability distributions for the deteriorated and non-deteriorated samples are, in most cases, recognized as significantly different. Obviously, all the major factors and most of the minor factors identified by RST are also recognized as related by MWU. The results obtained from the two approaches are almost identical for each type of deterioration, although a few discrepancies do exist. This is understandable since the notations as well as the algorithms of the two approaches are so different. For example, RST generates reducts by going through all attributes, while MWU adopts two sets of observations (i.e.,a factor vs. deterioration or not) for each run.

Table 12. Comparison of Results Obtained from RST and MWU

Factors	Cracking		Efflorescence		Corrosion of Rebar		Spalling		Breakage		Honeycombing	
	RST	MWU	RST	MWU	RST	MWU	RST	MWU	RST	MWU	RST	MWU
Bridge Age	□	○	□	○	□	○	□	○	□	○	□	×
No. of Spans	□	○	□	○	□	○	□	○	□	○	□	○
No. of Lanes	■	○	□	○	□	○	□	○	□	○	□	○
Length of Bridge	□	○	□	○	□	○	□	○	□	○	□	○
Area of Bridge Deck	□	○	□	○	□	×	□	×	□	○	□	○
Max. Span	□	○	□	×	×	×	×	×	□	○	□	○
Traffic Volume	□	○	□	○	■	○	□	○	■	○	□	○
Distance from Coast	□	×	□	×	□	○	□	○	□	×	□	○
Avg. Annual Rainfall	□	○	□	○	□	○	□	○	□	○	□	○
Avg. Rainy Days per Year	□	○	□	○	□	○	□	○	□	○	□	○
Peak Monthly Rainfall	■	○	■	○	■	○	■	○	■	○	■	○
Max. Rainy Days in a Month	■	○	■	○	□	○	□	○	■	○	■	○
Matching Accuracy	92%		83%		92%		92%		92%		92%	

■: major factor □: minor factor ○: related ×: not related

6. CONCLUSIONS AND SUGGESTIONS

6.1. Conclusions

In this study we first select 29 possible factors leading to RC bridge deck deterioration. RST is then used to match the factors to the corresponding deterioration type. To facilitate this process we use visual inspection data for 2,128 bridges, as well as the relative weather and traffic records. The gathered data are pre-processed by K-Means Clustering and Code Conversion. A brief review of the collected data shows that the distribution of the samples showing deterioration matches the sequence of deterioration types normally occurred in RC structures. The findings obtained as a result of data mining show that weather-related factors are rather significant in almost all types of deterioration. Furthermore, the short-term and intensive effects, such as peak monthly rainfall and the maximum number of rainy days in a month, have an even bigger effect on bridge deck deterioration. In addition we find that whether over water or not, number of lanes, type of pier, and type of girder material, designed live load and the type of expansion joint are major factors in cracking. Factors in cracking such as type of pier and type of girder material also turn up in efflorescence. It appears that cracking and efflorescence are somewhat correlated. "Traffic volume" is identified as a major factor leading to rebar corrosion and breakage. However, distance from coast is not significant to rebar corrosion in this study as our samples are considerable away from coast.

To validate the approach, the factors mined by RST are compared to the results obtained by MWU. It is found that the results are quite close although few discrepancies exist.

6.2. Suggestions for Future Research

Although this study demonstrates a systematic approach to identify the deterioration factors of RC bridge decks, some efforts can be made in future such as:

The MWU test requires observations to be in an ordinal or rating scale; categorical factors cannot be tested. An appropriate statistical approach is required to assess the correlation between categorical factors and their corresponding deterioration type.

Visual inspection data are often seen to be rather subjective and error-prone. It is recommended that steps be taken to eliminate logical mistakes or irrational records before performing the analysis.

A systematic approach for identifying RC bridge deck deterioration factors is introduced in this study. The same approach can be carried out for other bridge components. It is encouraged that the method be developed further. An expert system for diagnosis of bridge health based on the factors identified could be established.

ACKNOWLEDGMENTS

The authors wish to express their gratitude for the support of the work by the National Science Council, Taiwan under Grant No. NSC 95-2221-E-008-175.

REFERENCES

[1] Scherer W.T. & Glagola D.M. (1994), Markovian models for bridge maintenance management, *Journal of Transportation Engineering*, Vol.120, No.1.

[2] Zhao Z. & Chen C. (2002), A fuzzy system for concrete bridge damage diagnosis, *Computers and Structures*, Vol.80, pp.629-641.

[3] Su H.J. (2003) A correlation study of the existing bridges for failure analysis – case study of Taichung County, *Master Thesis at FCU*, Taichung, Taiwan.

[4] Chang J.W. (2004) A deterioration model of components of RC bridges in Taiwan, *Master Thesis at NCU*, Taoyuan, Taiwan.

[5] Huang R.Y. & Hsu W.Z. (2005) Reliability-based component deterioration model for bridge life-cycle cost analysis, *Journal of Chinese Institute of Civil and Hydraulic Engineering*, Vol.17, No.4, pp. 679-691

[6] Huang Y.H. (2003) Modelling deterioration of concrete bridge decks using neural networks, *PhD Thesis at UW*, Madison, Wisconsin, US.

[7] Chen T.Y. (2005) A study on the prediction of degradation of the freeway bridges by using neural networks, *Master Thesis at NTUT*, Taipei, Taiwan.

[8] Pawlak Z. (1982) Rough sets, International Journal of Computer and Information Science, Vol.11, No.5, pp.341-356.

[9] Wu C.Z. (2004) A research of data mining technology applied to marketing for property insurance, *Master Thesis at CYCU*, Taoyuan, Taiwan.

[10] Lin Y.W. (2005) Damage causality and maintenance strategies of asphalt concrete pavements, *Master Thesis at NTOU*, Keelung, Taiwan.

[11] Liu Y.S. (2005) Disease classification model based on rough set theory - using diagnosis of abdomen pain as an example, *Master Thesis at NYUST*, Yunlin, Taiwan.

[12] MacQueen J.B. (1967), Some methods for classification and analysis of multivariate observations, *Proceedings of the 5th Berkeley*

[13] *Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297.

[14] Ziarko W. & Ning S. (1995) Discovering attribute relationships, dependencies and rules by using rough sets, Proceedings of the 28th Annual Hawaii International Conference on System Sciences Vol 3, Issue 3-6, pp. 293 – 299.

[15] University of Warsaw (2005) RSES 2.2 User's Guide, http://logic.mimuw.edu.pl/~rses.