A Big Data Approach for Decision Making in Bridge Maintenance

Yu-Han Chuang and Nie-Jia Yau
Graduate Institute of Construction Engineering and Management, National Central University, Taoyuan, Taiwan
Email: yuhan@cc.ncu.edu.tw, yau@ncu.edu.tw

***Abstract*—The Taiwan Bridge Management System (TBMS) has been online since 2000 and the total amount of inventory is 33,275, including all kinds of bridges and culverts. Currently, the number of fields in all tables in the databases of TBMS is around 6,500 with more than 3.4 million data records in its databases. There are more than 11,200 bridges that are over 20 years old with another 9,300 bridge having unknown built years in the TBMS. The bridges in Taiwan have stepped into the stage where maintenance is crucial and frequently required. Therefore, this research aims at analyzing the database in the TBMS using Exploratory Factor Analysis for determining maintenance strategies for these bridges. This paper describes results of the first year's research efforts. Relevant literature in bridge maintenance, prioritization, and life-cycle bridge management were firstly reviewed. Concepts, theories, and available software for analyzing "Big Data" were also introduced.**

***Index Terms*—bridge maintenance, bridge management, big data, decision making**

I. Introduction

The Taiwan Bridge Management System (TBMS) has been online since 2000 [1]. Its inventory includes 28,365 bridges with 6,524 attributes in all tables, and the total amount of data records is 3,457,274 which increase 15,000 records annually. Among these attributes, there are 475 attributes containing kernel management information of a bridge such as inventory data, inspection results, and repair records. Thus, the databases in the TBMS have met the definition of "Big Data."

TABLE I. Bridge's Age of Each Management Agency

Bridge's age	Freeway/Highway	Railway	City/ County	Total
~10	726	54	1,235	2,015
10~20	2,262	262	3,198	5,722
20~30	1,004	332	4,037	5,373
30~40	971	359	2,752	4,082
40~50	252	408	606	1,266
50~60	61	104	199	364
60~70	18	65	41	124
70~80	9	9	9	27
80~90	6	-	15	21
90~100	-	-	5	5
100~110	-	-	1	1
unknown	16	135	9,214	9,365

Manuscript received July 11, 2015; revised January 12, 2016.

Age distribution of bridges in the TBMS is shown in Table I. There are 11,263 bridges over 20 years old, and Fig. 1 shows the amount of bridge components which are deemed necessary for maintenance actions. The number of seriously deteriorated components still increases gradually, even though maintenance activities have been expedited by responsible agencies for many of such components.

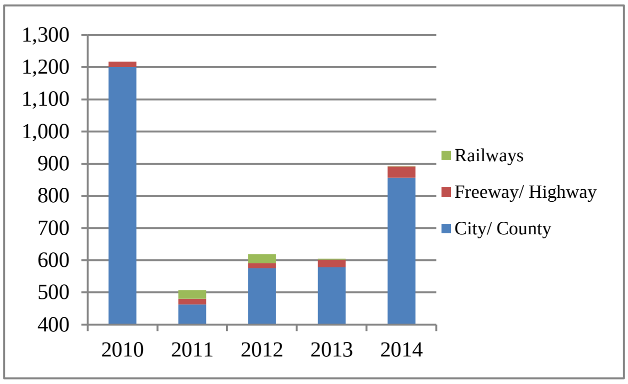

Figure 1. Amount of bridge components need maintenance actions

Due to limited budgets, especially for local governments, prioritization of bridge maintenance is always a tough task for the bridge management agencies, in addition to determining which option is better between maintaining and rebuilding of the bridge. Life-cycle cost analysis is a feasible solution for such problem; however, such technique requires an appropriate deterioration prediction model which does not yet implemented in the TBMS.

In order to effectively evaluate cost efficiency of repairing work and rebuilding of bridge, this research aims to analyze the TBMS databases to obtain characteristics of bridge deterioration in Taiwan that are useful for determining maintenance strategies. This two-year research project has two stages. For the first year, in addition to literature review in bridge maintenance, algorithm and available software related to big data are thoroughly investigated; and application of these techniques and software to the TBMS databases is planned to be performed in the second year.

For this research, the anticipated result is to obtain maintenance information such as repairing cost and period of components, progressive of deterioration conditions, and factors that trigger the repairing actions. Finally, a decision support and evaluation model for

doi: 10.18178/ijscer.5.3.216-222

rebuilding of deteriorated bridges will be established from this research.

II. Literature Review

A. Bridge Deterirarion Factors in Taiwan

Lin [2] successfully established a service life prediction model for expansion joint that obtained a 9% difference between the predicted and the actual service year. He also discovered that horizontal acceleration, number of spans and traffic flow are the most significant factors in determining the service life of an expansion joint. In addition, Su [3] collected 935 bridge inspection data in Taichung to analyze the relevancy between bridge deterioration and its environment by a logistic regression approach. This study discovered that the age of bridge, the distance to sea, and using of I-type girders are the major factors that caused deterioration.

B. Prioritization of Bridge Maintenance

For both the central and local governments, distribution of bridge maintenance budgets is always a difficult task. Chen [4] established a model to calculate a Danger Factor (DF) for a bridge by assigning weights to its major components based on their deterioration ratings multiplied by a traveler's factor determined by level of road that bridge was on; then the component having the highest value was normalized to represent the DF of the bridge. The DF can be used for both prioritization of bridge maintenance and distribution of maintenance budgets. This prioritization model is currently incorporated by the TBMS.

C. Effectiveness of Maintenance Budget

Budget spent for bridge maintenance needs to be effective. Feasible ways to check the effectiveness is to investigate results of maintenance within a time period or under limited budgets. Weng [5] compared the same amount of cost spent within a time period for fixing or replacing certain components to find which way is more effective. Huang [6] proposed a concept of concurrently maintaining multiple components on a bridge to reduce the overall time spent for repairing various components of the bridge. Lay [7] developed a maintenance cost analysis model that allowed the user to input the amount of budget for a given number of years, and the model would allocate the budget to the bridges to achieve the most effective result.

D. Bridge Life-Cycle Management

Many researchers have promoted life cycle cost concept for bridge management for many years. However, current practice in most bridge construction bids in Taiwan still not yet consider the maintenance costs. Zhu [8] established an optimal model for calculating bridge total life cycle cost for RC beam bridges, considering travelers' cost and social cost. Safi [9] analyzed the Sweden bridge management system to find a total maintenance cost for bridge components. The research results showed that the total maintenance cost is 15% to 25% of life cycle cost of a bridge, while different types of bridges may have more than 50% difference in construction cost.

E. Summary

Several studies in bridge deterioration factors and maintenance prioritization have established some analysis model for the bridges in Taiwan. However, actual maintenance frequency, costs, and methods of various bridge components could be used to generate a life cycle cost model which is crucial to obtain a more effective maintenance strategy. In addition, decision making between continuing maintenance actions and rebuilding of a new bridge still not clarified yet. Thus, answers to these doubts by digging into the actual inspection results and maintenance records in the TBMS have become the major objectives of this research.

III. Taiwan Bridge Management System (TBMS)

Supported by the Institute of transportation, Ministry of Transportations and Communications, the TBMS was developed by National Central University in 1999 in Taiwan. The TBMS is used by all the governmental agencies which are responsible for bridge management. In the TBMS, relational databases are incorporated in which relations of two or more data records are built by the same attribute. For example, if two databases, termed as "inventory" and "agency" are needed, an attribute named "agency's id" can be added into both databases to create a relation between them, as show in Fig. 2.

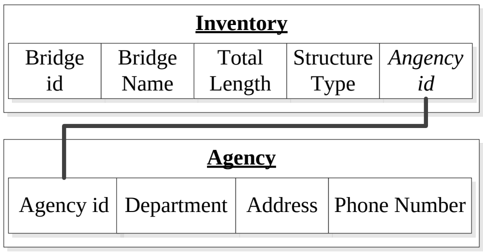

Figure 2. Data relation establishment

Fig. 3 illustrates the data structure of the TBMS. There are 30 data tables and 9 modules in the TBMS, such as Inventory, Inspection Data, Maintenance Records, Statistic, Decision Support, etc., as shown in Fig. 4. The Maintenance Codes in Taiwan regulates that every bridge needs to be inspected at least once per two years. Hence, this research focuses on data in three of these modules; they are Inventory, Inspection Data, and Maintenance Records modules, as described below.

A. Inventory Module

There are 33,275 data records in the inventory module; among which only 28,000 bridges are still in use or under maintenance, the rest were destroyed by natural disasters, closed or demolished due to serious deterioration. In this module, there are four tables that describe the basic data of a bridge. Bridge main inventory table is the top layer of data structure in this module; below which are abutment, pier, and span tables. The main inventory table consists of six kinds of data such as management, geometry, structure, particular structure, river, and design; the total number of fields is 147 with roughly 33,000 records accumulated since year 2000.

The abutment, pier, and span tables have data fields describing detailed geometry and design information with 42, 58, and 39 fields and around 9,700, 24,000, and 90,000 records, respectively.

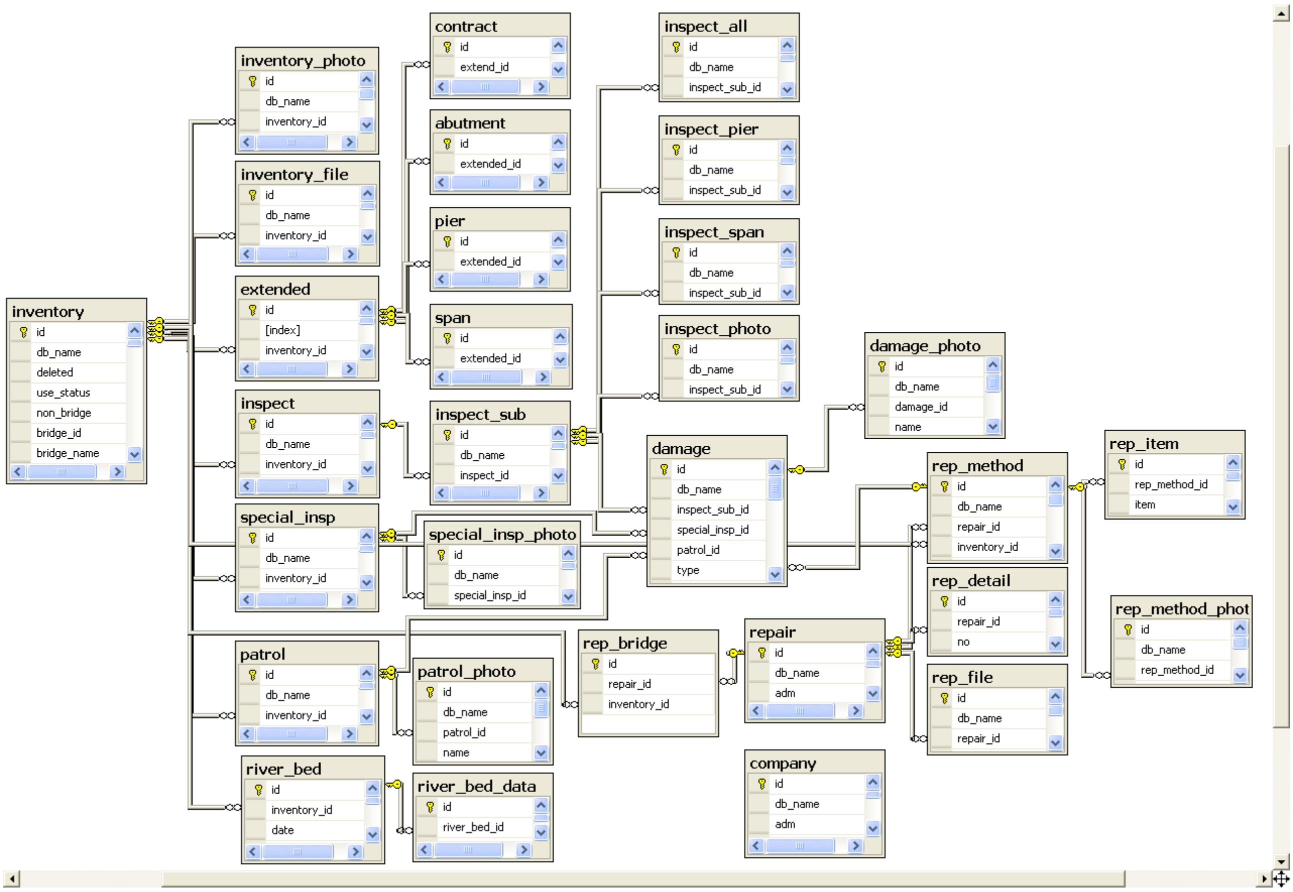

Figure 3. The diagram of database structure

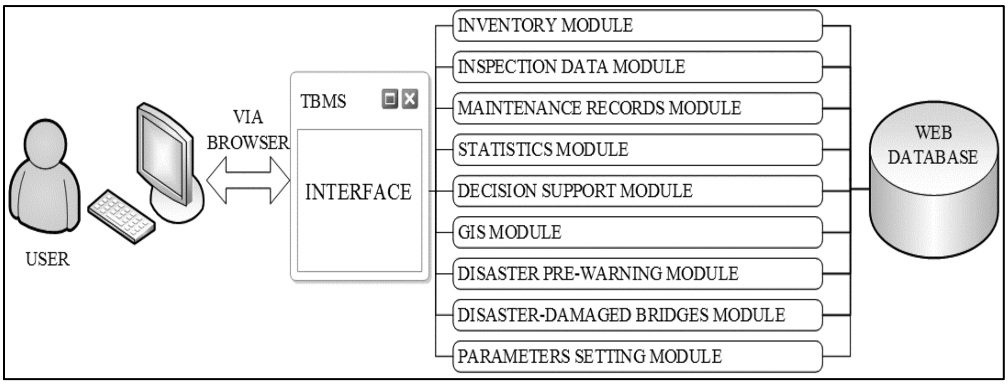

Figure 4. Major functional modules of the TBMS

B. *Inspection Data Module*

The methodology of regular bridge inspections used by the TBMS is called DER&U [10]. In this methodology, four indices are used to evaluate the condition of a bridge component: "D" represents the degree of deterioration; "E" represents the extent of the deterioration; "R" represents the deterioration's relevancy to bridge safety; and "U" represents the urgency for repairing the deterioration. All of these indices are numerically rated on an integer scale from 0 to 4 to describe the status of the deterioration, as exhibited in Table II. For a concrete bridge, 21 components need to be inspected, for other types of bridges the number of components may up to 25.

This inspection data module stores visual inspection results of all bridges. It has three layers of data structure; they are main, overall, spans and piers inspection sheets. These inspection sheets have 21, 69, and 51 fields to record the inspection results and currently they have around 276,000, 277,000, and 2,000,000 records, respectively. Since current regulation requires at least inspecting bridge once per two years, these records increase roughly 15,000 annually. Notably, if deterioration is found during inspection, it is required to input a suggested repairing method by the inspector. Thus, at the bottom of the data structure, the suggested repairing method is also recorded by 34 fields; it has 521,000 records in the TBMS now.

TABLE II. THE DER&U EVALUATION CRITERIA

	0	1	2	3	4
D	Component not existing	Good	Fair	Bad	Serious
E	Unable to inspect	Less than 10%	10~30 %	30~60 %	Over 60%
R	Relevancy uncertain	Minor	Limite d	Major	Large
U	Urgency uncertain	Routine	In 3 years	In 1 year	Immediately

C. Maintenance Records Module

In this module, there are seven tables used to record a maintenance work such as maintenance contract, contractor, and detail records of maintenance activities, etc. Currently, 54,000 maintenance records are stored in this module. The maintenance time, method used, costs and quantity of repaired components of a bridge are deemed as crucial information in this research.

IV. RESEARCH METHOD

Processes of this research along with related algorithms and software are described below. Attributes and tuples which will be analyzed are also depicted.

A. Research Processes

Fig. 2 shows the processes of this research. Notably, data of 9 tables will be firstly exported from the modules of inventory, inspection data, and maintenance records in the TBMS. The attributes used to establish relation of data will be deleted, and the repeated attributes will be excluded as well. After importing these data into a selected analysis software, the following results are expected to be obtained: (1) the major maintenance components, (2) the actual maintenance period of components, (3) the actual maintenance cost of components, and (4) the relation between inventory and Condition Index (CI). Coefficient of factors among attributes will be examined to make appropriate modifications during these processes.

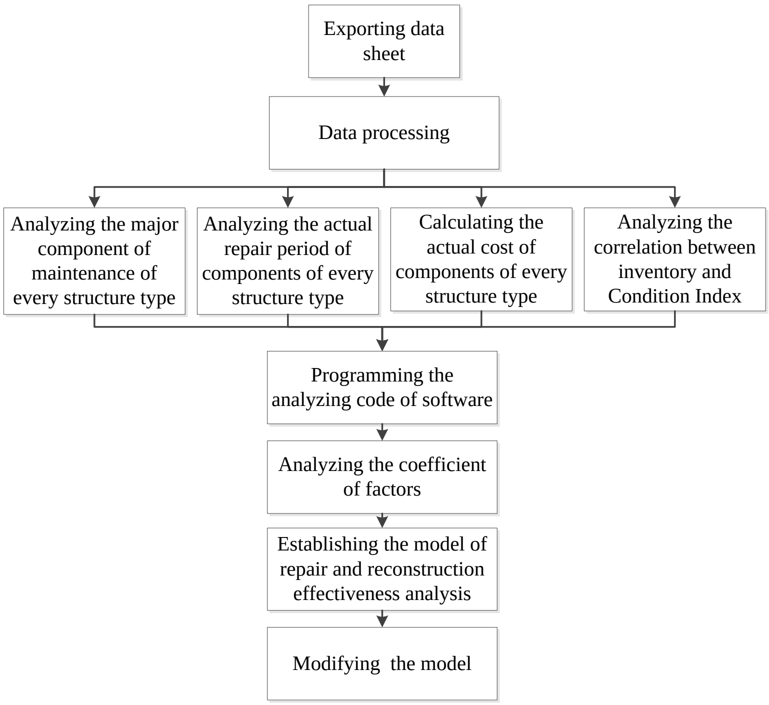

Figure 5. Diagram of research processes

1) Deleting the relational attributes

The TBMS is composed of relational databases; therefore there are many invalid attributes during data analysis. Fig. 3 is a part of repair record table where the top row shows the name of attributes. Column C is named repair_id, and column D is named inventory_id. These two fields show that this table is related to inventory and repair contract tables. These kinds of attribute are not necessary during data analysis and need to be deleted to decrease the amount of data.

2) Excluding lack of integrity tuples

There are about 3 million tuples in inventory, inspect data, and repair work. Some tuples might have null values in attributes. Fig. 4 demonstrates part of the pier data. The attributes named pylon_type, pylon material, and anchor_type are null in row 5, 6, and 7. Null values would make the result inaccurate, and they affects the result of factor analysis as well. So, the tuples with null values would be excluded in this research.

	A	B	C	D	E	F	G	H
1	id	db_name	repair_id	inventory_id	method	amount	unit	price
2	1	tanfb	1	17		2		210
3	2	tanfb	2	17		1		1253
4	3	tanfb	3	17		1		315
5	4	tanfb	4	40		5		525
6	5	tanfb	5	40		42		42882
7	6	tanfb	6	40		1		315
8	7	tanfb	7	40		1		315
9	8	tanfb	8	40		2		2506
10	9	tanfb	9	40		1		315
11	10	tanfb	10	40		1		1115
12	11	tanfb	11	53		1		0
13	12	tanfb	12	54		1		0
14	13	tanfb	13	56		3		384
15	14	tanfb	14	56		1		1287
16	15	tanfb	15	56		1		1287
17	16	tanfb	16	56		1		1287
18	17	tanfb	17	539		9		13414
19	18	tanfb	18	545		3		3424
20	19	tanfb	19	418		0.04		6.72
21	20	tanfb	20	418		0.12		20.16

Figure 6. Repair record table

	A	B	C	L	M	N	O	P	Q	R	
1	id	extended_id	no	pile_num	pylon	pylon_type	pylon_material	anchor_type	expansion	exp_cm	bearing
2	1	9675	1								
3	2	9509	1								
4	3	9509	2								
5	12	25991	1	14	否				無	0	簡易夾
6	13	25991	2	10	否				無	0	其他特
7	14	25991	3	10	否				齒型	16	盤式支
8	15	25991	4	24	是	單柱式	鋼構造	錨定式	無		合成橡
9	16	25991	5	24	是	單柱式	鋼構造	錨定式	齒型	32	合成橡
10	19	15179	1								
11	20	15179	2								
12	21	15179	3								
13	22	15179	4								
14	23	15179	5								
15	24	15179	6								
16	25	15179	7								
17	26	15179	8								
18	27	3688	1		否				齒型	0	合成橡
19	28	3688	2		否				齒型	0	合成橡
20	29	3688	3		否				齒型	0	合成橡
21	30	3678	1	0	否				齒型	0	合成橡
22	31	3678	2		否				齒型	10	合成橡
23	32	3678	3	0	否				齒型	0	合成橡

Figure 7. Pier data table

B. *Research Methodology*

Factor analysis is an algorithm that extracts latent variables among common observed variables to analyze the correlation between factors. This research aims to probe relation of attributes thus the exploratory factor analysis is incorporated. This methodology consists of 4 steps: extracting common factors, deciding the amounts of factors, rotating the axis, and naming factors; in which extraction and rotation are the most important ones as described below.

1) Extraction

Image factoring method calculates the image scores of variables through multiple regression, and it do principal component analysis by the covariance matrix of image scores. After calculation, it brings out a factor loading matrix which can be transformed into an eigenvalue matrix L. the mathematic model is shown in Eq. (1) and Eq. (2).

$$\mathrm{L} = \mathrm{VR}V' \tag{1}$$

$$\mathrm{R} = \mathrm{VL}V' = V\sqrt{L}\sqrt{L}V' = \left(V\sqrt{L}\right)\left(\sqrt{L}V'\right) = \Lambda\Lambda' \tag{2}$$

Where, L is eigenvalue matrix.

V is eigenvector.

Λ is factor loading matrix.

2) Rotation

The purpose of rotating factor axis is to clarify the relation of factors to reflect a simple factor structure. Since the definition of relational data model is independent of attributes, the orthogonal rotation method is deemed to be the appropriate way to process. This method makes the variance of factor loading square become maximum in order to simplify explanation of factors.

C. Big Data Analysis Software

Big data analysis has become a popular issue recently. After a thorough review of current available software, 11 kinds of popular software packages are found. They are Matlab, SAS, R, Python, Julia, Java, Hadoop and Hive, Scala, Kafka and Storm, Octave, and GO. Among which, Matlab and SAS are widely used by the academia, while R is incorporated by many famous portals. Thus, this research plans to utilize these three kinds of software packages to perform the big data analysis. Their characteristics are depicted below.

1) R

R was developed by Professors Ihaka and Gentleman at the University of Auckland in New Zealand. R is written for statistic, drawing, and data mining. R is capable of performing 25 kinds of statistic and numerical analysis functions such as obtaining mean value, standard deviation, plotting of histogram, and executing regression process. Most importantly, the source code of R is available freely. Its famous users include Google, Facebook, Bank of America, and New York Times.

In addition to the above functions, R can be used for matrix calculation; its efficient performance can be comparable to GNU Octave and Matlab. Thousands of added software tools based on various analysis techniques for economics and finance have been established on R by various languages such as LaTeX, JAVA, C, and FORTRAN.

3) Tatistics Analysis System (SAS)

Developed by SAS Institute Inc., SAS has been commonly used in commercial areas for decades (Wikipedia, 2014). The initial version of SAS was written in language C, and now JAVA and C++ are also included. Its latest version is 9.4, including 10 main modules for data mining, graphics and presentation, econometrics and time series analysis, clinical trial analysis, statistics analysis, interactive matrix language, quality control, and database transfer, etc.

2) Matlab

It is commercially available software developed by MathWorks. It can be used for algorithm generation, data visualization, data mining, data analysis and calculation. Its latest version is R204b which allows the user to establish user surfaces by its programing language or by calling other programs written by C, C++, JAVA, Python or FORTRAN.

Matlab also provides an easy-to-use tool box established based on various techniques such as factor analysis, generic algorithm, neural networks and ANN, allowing the use to perform functions such as optimal analysis, statistics, signal processing, image-processing, vector analysis, and matrix calculation. Notably, raw data preparation is crucial for Matlab since that may affect calculation efficiency.

D. Anticipated Result

The next step of this research taken is to formulate single data records which consists of data of fields from tables of bridge inventory, span, pier, abutment, main inspection, detail inspection, suggested maintenance method, and maintenance record. Data records have missing data in any field or have logic inconsistence will be eliminated. These records will be input to the three software packages; Matlab, SAS, and R as mentioned above. The anticipated results will be (1) a maintenance frequency for all the bridge components, (2) most maintained bridge components, (3) actual maintenance costs for bridge components, and (4) the relationship between CI and bridge inventory data. Finally, an evaluation model will be established for determining continuation of maintenance or rebuilding of a bridge based on these findings.

V. Conclusion

This research collected relevant literature in bridge maintenance and life-cycle costs analysis in Taiwan. It was found that models for calculating bridge life cycle costs still not yet established, nor the effectiveness comparison between maintenance and rebuilding of a deteriorated bridge. These have become goals of this research and are intended to be solved by digging into the big databases of the TBMS which has already been used for 15 years. This research depicts the preliminary steps in data processing to decrease the amount of data and increase the accuracy, and selects an appropriate algorithm and three available software packages for big data analysis. The data in the TBMS would be soon applied to find relevant maintenance information for decision making in bridge maintenance in Taiwan.

Acknowledgment

The authors appreciate the financial support provided by Ministry of Science and Technology, under project number MOST 104-2221-E-008-102-MY2.

References

[1] Institute of Transportation, MOTC, 2014. Taiwan Bridge Management System. [Online]. Available: http://tbms.iot.gov.tw/bms2/

[2] J. J. Lin, "Modeling prediction of service lives of bridge expansion joints," Master's thesis, National Yunlin University of Science and Technology, 2007.

[3] H. J. Su, "A correlation study of the existing bridges for failure analysis-case study of Taichung County," Master's thesis, Feng Chia University, 2003.

[4] J. J. Chen, "Developing a maintenance decision support module for Taiwan bridge management system– An example for directorate general of highways," Master's thesis, National Central University, 2007.

[5] K. H. Weng, "Comparison of economic efficiency of rehabilitation and replacement in bridge maintenance," Master's thesis, National Central University, 2009.

[6] H. Y. Huang, "Establishment of bridge elements concurrent maintenance model," Master's thesis, National Yunlin University of Science and Technology, 2007.

[7] Y. C. Lay, "A maintenance strategy evaluation model for network level bridges," Master's thesis, National Central University, 2001.

[8] J. Zhu and B. Liu, "Performance life cost-based maintenance strategy optimization for reinforced concrete girder bridges," *Journal of Bridge Engineering*, vol. 18, no. 2, pp. 172-178, 2013.

[9] M. Safi, H. Sundquist, and R. Karoumi, "Cost-efficient procurement of bridge infrastructures by incorporating life-cycle

cost analysis with bridge management systems," *Journal of Bridge Engineering*., vol. 20, no. 6, p. 04014083, 2014.
[10] Institute of Transportation, MOTC, Establishment of Bridge Visual Inspection and Evaluation Manual (Draft), 2011.

Yu-Han Chuang was born in Taiwan, 1986. She is a Ph.D. student of Graduate Institute of Construction Engineering and Management, National Central University, Taiwan.

Nie-Jia Yau was born in Taiwan, 1958. He obtained his Ph.D. degree from Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign in 1992. He is now Professor of Graduate Institute of Construction Engineering and Management and also Director of Research Center for Hazard Mitigation and Prevention, National Central University, Taiwan.