A Framework for Developing an Estimation Model of Damages on Bridge Elements Using Big Data Analytics

S. Lim[a*], S. Chung[a], S. Chi[a], and J. Song[a]

[a]Department of Civil and Environmental Engineering, Seoul National University, Republic of Korea

E-mail: sorami@snu.ac.kr*, hwani751@snu.ac.kr, shchi@snu.ac.kr, junhosong@snu.ac.kr

Abstract

According to the Ministry of Land, Infrastructure and Transport in Korea, the number of bridges over 30 years old will be three times in 2025 than three thousands in 2015 so that proper maintenance enforcement has emphasized. The Korean government legislated on the Special Act on Safety Control for Infrastructure in 1995 and rated the condition grade of "A" to "E" to indicate that under the "C" grade of infrastructure requires critical repairs. Since the lack of maintenance budget and decrease of professional inspectors, more effective and efficient monitoring solutions for bridge conditions are required. The primary purpose of this study is to develop a model to predict deficiencies of bridge elements using big data analytics and the framework is proposed in this paper. The model analyzed a dataset of Bridge Management System (BMS) developed by the Korea Institute of Civil Engineering and Building Technology (KICT) (e.g., general bridge information, structural information, and inspection and maintenance history) and found combinations of significant factors causing damages of bridge elements. Data mining algorithms (Apriori and Ripper) were applied for the pattern analysis and the partitional clustering algorithm grouped similar patterns for more efficient and fast damage prediction. Machine learning concepts using bagging and boosting were also employed for resampling and upgrading the estimation model to enhance estimation performance. The expected results showed potential to predict causes, locations, current status, and the types of bridge damages that can be used for preventive bridge maintenance planning.

Keywords

Bridge Management; Infrastructure Maintenance; Damage Estimation; Big Data

1 Introduction

According to the World Economic Forum, infrastructure plays a key role for economic development since it can enhance national competitiveness. In addition, the condition of infrastructure is crucial because its deterioration can threat public safety and hinder economic activity[1]. Recently it is reminded owing to a case that a serious damage of steel wire to support main structure of the elevated bridge of Jeongneung stream in Seoul, Korea was found by emergency inspection and then the bridge was blocked for a month from February 2016. Bridge structures need enough time and budget for maintenance inspection considering its structural complexity and also direct and indirect impacts of deficiency, and thus preventive management have become significant.

A number of bridges were built with rapid economic growth of Korea in the 1970s and then the number of bridges over 30 years old is expected to rise over three times in 2025 from 3,094 in 2015[2]. Nowadays many of them have deteriorated and required proper maintenance[3]. After the collapse of Seongsu Bridge in Seoul in 1994, the Korean government legislated on the Special Act on Safety Control for Infrastructure in 1995 to designate major facilities as type Ⅰ and type Ⅱ facilities and manage them. The aim of the Act is to provide methods and regulations of periodic inspection for proper repairs based on infrastructures' condition grade of "A" to "E". A bridge under the state "C" requires critical maintenances.

Current bridge management systems in Korea only consider about ten thousand major targeted bridges based on its scale elements. Type Ⅰ bridges cover the bridges which have 1) a specific superstructure types including suspension bridges, cable-stayed bridges, arch bridges, and truss bridges, 2) maximum span length over 50m except one span bridge, 3) length over 500m, and 4) cut and cover structure with width over 12m and length over

500m. Type Ⅱ bridges are the second priority structures, and they include the bridges which are 1) one span bridges with span length over 50m, 2) length over 100m among not the type Ⅰ bridges, 3) cut and cover structure with width over 6m and length over 100m among not the type Ⅰ bridges. However, approximately twenty thousand smaller bridges which are under 100m in length, not targeted to be managed by the Act, are exposed to be in danger. Nonprofessional inspectors have been hired and even such an inspection has not been conducted regularly[4].

An uniform inspection cycle based on the condition grade also hinder efficient investigator arrangement and budget allocation for weak facilities. Since there are various causes of bridge damage and the time of occurrence are less predicted[5], there is a need for a model to support inspections at ordinary times.

Developing a model to support effective and efficient bridge inspection had been a critical issue in bridge management so many studies have been conducted not only in Korea, but also in many other countries. The bridge condition deterioration model have been one of main streams providing condition ratings. The models can be categorized into three groups which are not mutually exclusive: deterministic, stochastic, and artificial intelligence[6].

Deterministic models including regression models represent the relationship between the factors (e.g., bridge age, length, and width) that affect deteriorations and the facility functions based on mathematical or a statistical formulations. Stochastic models such as Markovian models focus on facility condition change from one state to another during one inspection period using probabilities of one or more random variable. Artificial intelligence (AI) models with data mining techniques comprise classification, clustering, and mining frequent patterns. Classification is to find a model that explains and distinguishes data classes to predict the class label of objects for which the class label is unknown. Artificial Neural Network have been utilized in many studies to predict condition ratings of bridge[7-10]. Clustering forms some groups based on similar characteristics measured by distance between data points. The latter analysis is advantageous to find frequent patterns which are a set of items that often appear together[6], [10-13]. Huang and chen (2012) clustered bridge data in similar characteristics and found mining association rules using the National Bridge Inventory in the United States[14], but this research did not show the relationships with bridge deficiency and just described the associations between conditions.

Regression models are simple but they have several constraint conditions. At first, they need to select several factors influencing bridge condition grade before making models and thus some factors can be ignored even if they are significant[15]. Second, the models are updated as calculating new coefficient, and therefore, it is hard to update frequently. Moreover, in the context of application, most regression studies have focused on the condition grade of the whole bridge rather than its elements or different types of elements' deterioration because a parsimonious model, a model that satisfies a desired level of explanation or prediction with as few independent variables as possible, is usually preferred[16]. Even though deterioration factors that cause bridge elements' problems were not much different by each type of elements and thus individual properties of elements did not properly reflected. For instance, the current deterioration model in the BMS uses combinations of factors among bridge age, span length, Average Daily Truck Traffic, average humidity, and amount of surface chloride as independent variables for different type of elements. Since the number of variables is small and the combinations are not much different by type of elements, the current model did not very much good at considering different type of elements[17].

Stochastic models such as Marcovian model are not suitable to develop a model to assist inspections since they have been applied for prediction of future conditions of bridges, not correspond with the objective of this research focusing on present condition predictions. Moreover, the Marcovian model based only on the current condition without considering historical records which make this model unrealistic according to [6], [18].

In this context, adopting the AI approach using big data analytics or data mining techniques to discover patterns is suitable for developing a model. Big data analytics enables the model to use as many as variables for more reliable model development. Easy updating of algorithms enhance the estimation performance and also can encourage frequent uses. Inspections supported by such functions will be more efficient by offering the elements which have high probability of deficiency.

The primary purpose of this study is to develop an easily updated estimation model for bridge elements' damage by considering numerous factors from the bridge element level with big data analytics. The model is to find a set of patterns such as sheet waterproofing as a deck waterproofing type and the maximum span length between 29.3 and 32.3 that can impact to each element's condition grade from A to E (e.g., deck delamination = "D").

This paper developed a model by using Bridge Management System (BMS) data collected by the Korea Institute of Civil Engineering and Building Technology (KICT). There are two types of bridge condition ratings in BMS which are a state condition grade and safety condition grade. The latter is assessed by structural analysis and load carrying capacity test and therefore, it is hard to be replaced. The estimation object is state condition grade and it will be mentioned as just “condition grade”. BMS data covers bridges categorized as general national road, and smaller bridges as well as type Ⅰ and type Ⅱ bridges.

The word “prediction” in this research implies that the model estimates probable damages in target bridge elements. The prediction does not mean time series analysis.

This paper introduces the framework to develop the estimation model and preliminary results.

2 Research Methodology

2.1 Research Framework

The major steps to build the estimation model were described in Figure 1.

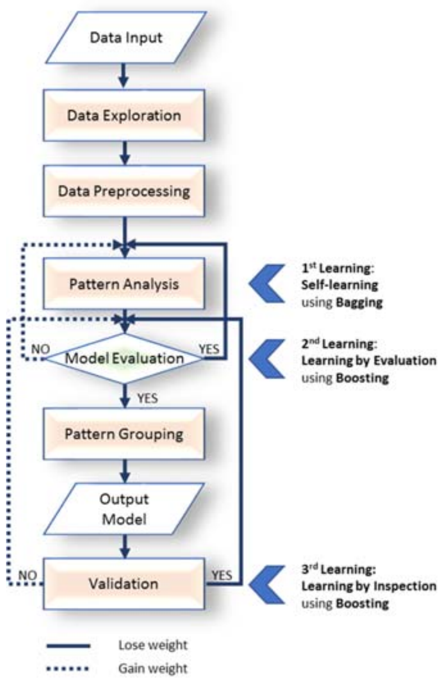

Figure 1. Research Framework

First of all, data exploration to understand data characteristics was performed. Data types (e.g., numerical variable and text variable) and data distributions were identified in this step. The next step was data preprocessing which covered data reduction, data transformation, data discretization, data integration, and data cleaning. Using the preprocessed data, pattern analysis including associations and classifications was carried out and this paper presents the preliminary results by this step. As one of machine learning algorithms, self-learning algorithm such as bagging, and an algorithm learning by evaluation such as boosting are applied to improve estimation performance. Significant patterns are then grouped by the similar input conditions or the same output damage. At the final stage, the model is validated using new inspection data by the boosting algorithm. All procedures were carried out in R software version 3.2.2 for statistical computing and graphics.

2.2 Data Characteristics

Accumulated inspection data in BMS were categorized into three parts: general factors (e.g., bridge class, locations, competent authorities, construction related factors, mileage, offset, detour, traffic volume, length, width, number of lanes, number of spans, main structure type, substructure type, design live load, and attached facilities), structural factors (e.g., span length, decks, girders, diaphragms, ribs for span and support types, abutments, piers, expansion joints, shoes, and stopper factors for support) and inspection factors (e.g., span or support number, element code, damage code, condition grade and maintenance record). Some text options were coded into numbers such as 11 indicated a RCS type of main structure. Factors were divided into discrete numerical variables (e.g. bridge class and main structure type), continuous numerical variables (e.g., mileage and traffic volume), and text variables (e.g., design firm, constructor, supervisor, and inspection descriptions). To utilize text string variables, they should be preprocessed into discrete values and they were not considered yet in this paper.

The size of BMS data were quite large enough to be called big data since it contained about 84 thousands tuples (i.e., records or rows). General data covered total 6,773 bridges and structural data included 19,625 rows for span and 25,729 rows for support. Inspection data was 10,655 which comprised 9,775 of detailed inspection and 900 of precise safety diagnosis. Detailed inspection data was used since it contained the condition grade in an element level. 834,815 rows inspection data in a level of different damage types of elements were composed of 3,701 bridges’ span data and 3,000 bridges’ support data.

The classes of 6,773 bridges in the data were type Ⅰ bridge (13.6%), type Ⅱ bridge (18.6%), the others (64.7%), and missing value (3.1%). A distribution of condition grade of the whole bridge showed grade “A” (27.9%) and “B” (64.7%) took large portion and “C” (3.6%), “D” (0.1%), “E” (0%), and the other values (3.7%). In the bridge element level, distribution is different from that of condition grade of the whole bridge:

the portion of grade "A" (34.4%) and "B" (37.0%) are smaller and "C" (11.2%), "D" (1.3%), "E" (0.1%), and the other values (16.0%).

2.3 Data Preprocessing

Variables used in the developed model consisted of predictors (i.e., input variables or independent variables) and targets (i.e., output variables or dependent variables), and preprocessing step was different according to the variable type and the role of each variables.

2.3.1 Predictors

Data reduction, data transformation, and data discretization were conducted for predictors. Predictors covered all general factors, all structural factors, and the condition grade and the maintenance record from inspection factors in order to consider condition grade. Data transformation was then performed to correct or erase typing errors and to have consistency of expression. For example, bridge elements' condition grade should comprised from "A" to "E" but number (e.g., 1, 2, and 3), some symbols (e.g.,] and '), and other characters (e.g., N, Q, V, and X) were sometimes included in a original dataset. Numbers and the other values were removed and small letters from "a" to "e" was changed into the capital letters.

Discretization was also carried out to make continuous numerical variable be categorized. It can be considered as if one dimensional partitioning and then a k-means clustering was utilized. By k-means clustering, the objects are distributed into k clusters based on a rule that minimizing within-cluster distances and maximizing inter-cluster distances[6]. The "discretize" function in the "arules" package was applied to make results on R software. To set the number of classes (k) before clustering, Sturge's rule was applied using size of sample (n) [19]. The calculated values were 20 for span and 19 for support when n was 560,272 for span and 240,418 for support.

$$k = 1 + 3.322(log_{10}\mathrm{n}) \qquad (1)$$

2.3.2 Targets

The major goal of this research is to find patterns of bridge elements' damages, so the related variables were restricted to represent target variables. An aggregated factor (e.g., 2B07100C) consisted of span or support number (2), element code (B07; expansion joint), damage code (100; crack), and condition grade (C) from inspection factors.

2.3.3 Generating Dataset

The preprocessed variables were combined by the bridge number which existed in the every data table. Span dataset (87 variables, 452,361 records) and support dataset (69 variables, 193,499 records) were generated to include not only predictors but also targets. Since there was not any common structural factor between span factors and support factors and thus two divided dataset were created.

2.3.4 Data cleaning

Handling missing values was critical because blank cells were randomly but frequently distributed in each dataset. There are two common approaches to deal with missing values: ignoring the tuple or filling in the missing values[11]. The former is very conservative and then the span dataset were shrunk as 66,542 tuples when it was applied. A global constant (e.g. 9999) or using a measure of central tendency (e.g., mean for numerical continuous variable and mode for numerical discrete variable) to fill in the missing value were used to find the best method that represented real population.

2.4 Data Mining

The core part of the estimation model was the pattern analysis and it had two axis in this research: descriptive mining and predictive mining. Descriptive mining is to comprehend essential characteristics of the data and answers to "What has happened?" and predictive mining deduces patterns from the data and replies to "What could happen?"[20], [21]. Two approaches used different techniques but the objective is same: finding patterns between predictors and targets.

The first descriptive axis was to find frequent patterns, which were association rules. The rule indicated only co-occurrence and did not mean causality. This process found significant factors that correlated with each bridge elements' damage. Association rules are sets of items (i.e. itemsets) which consist of "if" (i.e., antecedent) and "then" (i.e., consequent). Items indicate each condition such as bridge age under 20 years and a "frequent itemset" means the itemset can be found in the data more than a specific level called minimum support. "Support" is calculated as the number of occurrences with both antecedent and consequent itemsets divided by the number of the whole tuples. Another measure that shows more strong co-occurrences than support using conditional probability concept is "confidence". The "confidence" is calculated as support divided by the number of occurrences with only antecedent itemsets. It means how many occurrences which include consequents among the occurrences only with antecedents. The validity of the created rules is obtained by "lift" calculated as confidence divided by benchmark confidence. The benchmark confidence is computed by the number of occurrence with only consequent itemsets divided by the number of the whole tuples. Assume that a rule can be created as if six years old bridge and number of shoes under three then B16100C (i.e., pier, network of fine cracks, and grade "C"). In this case, benchmark confidence presents the frequency of B07100C. If lift ratio is smaller than 1, the obtained rule is not significant since there is a good possibility that B07100C can occur without antecedents [11], [21].

To mine association rules, Apriori algorithm proposed by Agrawal and Srikant in 1994[22] were used in this research since its efficiency for analyzing huge amounts of data. The basic concept of Apriori is that if an itemset is not frequent, each superset of the itemset does also not appear frequently. For instance, if {condition 1} → {condition grade C} is not frequent, then {condition 1, condition 2} → {condition grade C} is also not frequent. The contraposition for frequent cases is also true[11]. This concept prunes insignificant rules and then enhances the computational efficiency. Data discretization is needed to operate the Apriori algorithm. On R software, the "apriori" function in the "arules" package was utilized. Rules are created by adding conditions one by one until the number of conditions reaches the maximum length. The results contained a lot of combinations, and thus to obtain the desired outcome such as {condition 1, condition 2} → {condition grade C} an option was added that descendants should include the target variable column.

In the other axis, classification as performed to find patterns which combinations of certain conditions predict a particular condition grade. As one of the classification method, rule-based classification also finds rules (i.e., patterns) like mining association rules but its objective is to predict class label such as condition grade. For extracting predictive patterns, RIPPER (Repeated Incremental Pruning to Produce Error Reduction) as a rule producer[23] was applied. RIPPER directly extract rules from data without going through other classification models such as decision tree, so the computation time is faster.

In the RIPPER, one condition is created and the next condition is added with examination of how well the training data (i.e., data used for developing model) is explained. After making patterns, it sets a different training set from the whole data and compares rules from training sets. Through this step, the optimized combinations can be created[11], [24]. The "JRip" function in the "RWeka" package on R software was used.

After finding patterns, a partitional clustering algorithm groups similar patterns not only for efficiency but also for utility for users. The grouping includes two directions: one is to make clusters with common combinations of conditions, the other is to group using some component of the target such as condition grade, damage code, element code, and span or support number.

2.5 Updating

The developed estimation model is updated using ensemble algorithms of machine learning to obtain better prediction performance. Ensemble methods apply multiple classifiers which indicate an algorithms or a specific mathematical functions that implement classification[11].

In the pattern analysis phase, bagging, also called bootstrap aggregating proposed by Leo Breiman(1994) is utilized to reduce variance and to avoid overfitting of classification. The main concept of this method is creating multiple training sets by random sampling from the given training set, building estimation models in each sample space (i.e., bootstrap), and eventually combining estimation models using majority vote for classification[11], [24].

In addition, boosting algorithm is applied for model evaluation and validation. To decrease misclassification and maximize the classification performance, misclassified items gain weight and correctly classified items lose weight during boosting so that weak classifiers concentrate more on the items previously misclassified[11,] [24]. Model update occurs when a new data enters. Training population is changed and then bagging algorithm builds estimation model.

2.6 Validation

The developed model is validated by comparing new inspection data to the estimation results. After entering the data about newly inspected bridge into the model, the model matches the data to patterns found and evaluate whether inspected condition grade corresponds the predicted grade. The validation result will be an input data for boosting and the classifiers will be weighted.

3 Results

3.1 Preliminary Results

The 11 rules and 678 rules were found using the Apriori algorithm in the span and support dataset respectively with numerical settings that 0.001 for minimum support, 0.5 for minimum confidence, and four for maximum length of itemsets (Table 1). Four patterns from the support dataset were created using the RIPPER without any numerical settings. The number of patterns including only grade under C were selected since the bridge elements of condition grade under C were the targets to get proper maintenance.

Table 1. Obtained patterns of bridge span based on the Apriori algoritm

Combination of conditions			Estimated condition grade
Condition 1	Condition 2	Condition 3	
Inspection date =20071016	Maximum span length =[29.3,32.3)	City =Moon-Kyung	Expansion joint
		Road code =2003	Crack
		Millage =[274.64,316.48)	C
Inspection date =20080808	Type of cross =Normal (not elevated)	Effective width =[18.36,20.83)	
		Deck main rebar space =[27.51,37.21)	
		Girder space =[207.2,229.6)	Deck
	Allowable load=0	Substructure type =Round-shaped pier	Crack
		Effective width =[18.36,20.83)	C
	Total width =[18.91,20.11)		
	Barrier existence =Y		

The results can explain that some structural conditions determines specific grade and type of bridge elements. The results about span using the Apriori shows that specific location, time, and structural characteristics had an impact for C grade of crack on expansion joint. When a bridge which was located on the load code 2003 in Moon-Kyung city with maximum span length between 29.3 and 32.3 and millage between 274.64 and 316.48 was inspected in 2007 it frequently showed C grade of crack on expansion joint. For C grade of crack on deck was influenced by time, structural types, and structural sizes. A bridge which was not the elevation bridge and had barriers and rounded-shaped piers with widths between 18.36 and 20.83, rebar space between 27.51 and 37.21, or girder space between 207.2 and 229.6 frequently represented C grade of crack on deck. The support data presented similar format results.

Using the RIPPER algorithm patterns about not span but support were found as follows:

1) If bridge Age=21 and abutment high=[0, 1.41) then grade D of crack on bridge bearing.
2) If minimum distance=[21.21, 27.33) and bridge age = 16 then grade C of erosion on bridge bearing.
3) If bridge length = [43.2, 78.2) then grade C of erosion on bridge bearing.
4) If bridge age = 24 and number of up-lanes = 2 then grade C of deformation on bridge bearing

The patterns from the RIPPER algorithms were simpler than that from the Apriori but the explanation power was not good enough since the number of conditions was too short and it is less understandable such as only bridge age and the number of up-lanes leaded to bridge bearings' erosion.

3.2 Expected Outcomes

To provide interpretable information the output is designed as a portfolio type of each bridge span or support as illustrated in Table 2. Using confidence probabilities of each pattern will also be offered.

Table 2. Sample portfolio of a bridge span

1) Bridge Number: 03447
2) Bridge Name: Songgang
3) Span number: 3

Combination of conditions	Element	Damage	G	Prob.
{bridge length =(970.0, 1289.3), average traffic volume =(18291, 22064), deck pavement thickness =(7.75, 8.50)}	Deck	Crack	C	70%
{bridge width =(17.85, 20.08), deck depth =(23.7, 27.5)}	Girder	Crack	D	60%
{no. of lanes=4, girder type =PSCB}	Diaphra-gm	Exposed Steel bars	B	80%

4 Conclusions

This study proposed a framework to develop an estimation model to find patterns of bridge element's deficiencies by applying association, classification, and clustering in big data analytics. The preliminary results implied that specific conditions of factors can cause bridge elements' damages. The developed model uses the number of various factors and can be updated using machine learning algorithms which are strengths of this model compared to existing regression models.

The study used BMS data which covers not only type Ⅰ and Ⅱ bridges but also smaller bridges and therefore the model can be utilized widely. Identifying many factors related with specific condition grade of each bridge element provides useful information to be referred for inspection that enables effective and efficient monitoring. In short, this estimation model can enhance current inspection system by enabling a preventive bridge maintenance planning.

This paper suggested the framework to develop the estimation model and provide preliminary findings of rules. Other association rule algorithms and classification algorithms should be compared and conducted to find optimized level of performance as well as remaining steps from the research framework. The detailed results, portfolio and a set of critical factors influencing bridge elements' damage, will covered in our future research.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) Grant (No. 2015R1A5A7037372) funded by the Korean Government (MSIP) and the Promising-Pioneering Researcher Program through Seoul National University(SNU) in 2015.

References

[1] ASCE. 2013 Report Card for America's Infrastructure, 2013.

[2] Ministry of Land, Infrastructure, and Transport of Korea. *Yearbook of road bridges and tunnels*, 2015.

[3] Lee Y. H. Construction market prospects related to improvement of infrastructure safety, 2014.

[4] Seok I.H., Gong S. C. and Jung S. H. A consideration of safety inspection for small scale vulnerable facilities. *The journal of Infrastructure Safety*, 37:7-26, 2011.

[5] Lee Y. S. The improvement of precision safety diagnosis cycle for bridge. *The Magazine of the Korean Society of Civil Engineers*, 49(6):62-72, 2001.

[6] Morcous G., Rivard H. and Hanna A. Modeling bridge deterioration using case-based reasoning. *Journal of Infrastructure Systems*, 8(3):86-95, 2002.

[7] Li, Z. B., Shi, Z. Q., and Ososanya, E. T. Evaluation of bridge conditions using artificial neural networks. In *Proceedings of IEEE Southeast Conference 1996*, pages 366–369, Tampa, Florida, United States, 1996.

[8] Kawamura, K., and Miyamoto, A. Condition state evaluation of existing reinforced concrete bridges using neuro-fuzzy hybrid system. *Computers & structures*, 81(18):1931–1940, 2003.

[9] Elhag T. and Wang Y. Risk assessment for bridge maintenance projects: neural networks versus regression techniques. *Journal of Computing in Civil Engineering*, 21(6):402-409, 2007.

[10] Callow D., Lee J., Blumenstein M., Guan H. and Loo Y.-C. Development of hybrid optimisation method for Artificial Intelligence based bridge deterioration model — Feasibility study. *Automation in Construction*, 31:83-91, 2013.

[11] Han J., Kamber M. and Pei J. *Data mining: concepts and techniques*, Third Edition. Elsevier, Waltham, United States, 2012.

[12] Jiang Y., Saito M. and Sinha K. C. Bridge performance prediction model using the markov chain. *Transportation Research Record*, 1180: 25–32, 1988.

[13] Kobayashi K., Kaito K. and Lethanh N. Deterioration forecasting model with multistage weibull hazard functions. *Journal of Infrastructure Systems*, 16(4):282-291, 2010.

[14] Huang, R. Y. and Chen, P. F. Analysis of influential factors and association rules for bridge deck deterioration with utilization of national bridge inventory. *Journal of Marine Science and Technology*, 20(3), 336-344, 2012.

[15] Huang R.-Y., Mao I. S. and Lee H.-K. Exploring the deterioration factors of RC bridge decks: a rough set approach. *Computer-Aided Civil and Infrastructure Engineering*, 25(7):517-529, 2010.

[16] Forster M. Parsimony and simplicity. Available from: http://philosophy. wisc. edu/forster/220/simplicity. html, 1998.

[17] The Korea Institute of Civil Engineering and Building Technology. Development of bridge cost and performance database and module, 2012.

[18] Madanat S. M., Karlaftis M. G., and McCarthy, P. S. Probabilistic infrastructure deterioration models with panel data. *Journal of infrastructure systems*, 3(1): 4-9, 1997.

[19] Nie N. H., Hull C. H., Jenkins J. G., Steinbrenner K. and Bent D. H. *SPSS statistical package for the social sciences*, Second Edition. McGraw-Hill, New York, United States, 1975.

[20] IBM Descriptive, predictive, prescriptive: transforming asset and facilities management with analytics. On-line: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=TIW14162USEN, Accessed 08/03/2016.

[21] Slimani T. and Lazzez A. Efficient analysis of pattern and association rule mining approaches.

arXiv preprint arXiv:*1402.2892*, 2014.

[22] Agrawal R. and Srikant R. Fast algorithms for mining association rules. *Proc. 20th int. conf. very large data bases, VLDB*, 1215:487-499, 1994.

[23] Cohen, W. Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning*, pages 115-123, Tahoe City, California, Unites States, 1995.

[24] Shmueli, G., Patel, N. R., & Bruce, P. C. *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. John Wiley and Sons, 2011.