62 Sequential Pattern Analyses of Damages on Bridge Elements for Preventive Maintenance

Kowoon Chang, Soram Lim, Seokho Chi,
and Bon-Gang Hwang

Abstract

For the safety and serviceability of aging bridges, it is important to understand how the current conditions of the bridges will deteriorate in the future as time goes by. The primary goal of this research is to analyze sequential patterns of damages on the bridge elements that are normally recorded through site inspections and managed by the bridge management system (BMS). To achieve the research goal, the research team first discovered a number of bridge clusters with distinguished characteristics by using a data clustering algorithm. Sequential pattern mining was then utilized to extract types and sequences of damages on bridge elements frequently seen in each cluster. The data used for the analyses was collected from BMS managed by the Korea Institute of Civil Engineering and Building Technology. This BMS includes the general, structural, traffic, and weather information of 6773 bridges (i.e., the total of 127 attributes) and contains 834,815 site inspection records of the bridge elements. A preliminary test was performed by using a dataset of 1542 Pre-Stressed Concrete I-shape type bridges for the validation purpose. The results of this study showed application potential to estimating future condition changes of the bridges based on the past inspection records for preventive bridge maintenance.

Keywords

Bridge management system (BMS) • Preventive maintenance • Big data analytics • Damage patterns • Sequential pattern mining

62.1 Introduction

In recent years, the number of aging bridges has rapidly increased all over the world, including South Korea. Almost 40% of bridges in the U.S. became older than 50-year-old in 2017 [1] and the number of aging bridges over 20 in South Korea is expected to increase up to three times higher in 2026 [2]. To ensure the safety and serviceability of the aging bridges, it is important to understand how the current structural and visual conditions of the bridges will deteriorate in the future as time goes by.

In response to the importance, the bridge management system (BMS) has been developed as a strategic tool in order to forecast future bridge conditions for better planning of maintenance, rehabilitation, and replacement. BMSs have established

K. Chang · S. Lim · S. Chi (✉)
Seoul National University, 1 Gwanak-Ro, Seoul, 08826, South Korea
e-mail: shchi@snu.ac.kr

K. Chang
e-mail: jgwoon1838@snu.ac.kr

S. Lim
e-mail: sorami@snu.ac.kr

B.-G. Hwang
National University of Singapore, 21 Lower Kent Ridge Road, Singapore, Singapore
e-mail: bdghbg@nus.edu.sg

I. Mutis and T. Hartmann (eds.), *Advances in Informatics and Computing in Civil and Construction Engineering*,
https://doi.org/10.1007/978-3-030-00220-6_62

traditional deterioration models that explain condition rating changes of bridge elements, such as individual decks of the superstructure, by using multi-regression or Markovian transition probabilities with significant deterioration parameters (e.g., bridge age, deck material, deck length, and ADT) [3–5]. For more reliable deterioration prediction, the applications of big data analytics into bridge maintenance have been proposed to discover meaningful knowledge from a big size of bridge data. Kim et al. [6] examined the massive National Bridge Inventory (NBI) dataset to identify the deterioration trends and develop a corresponding model in the U.S. by utilizing both deterministic and probabilistic analyses including two-factor analysis of variance. Similarly from the NBI database, Huang and Chen [7] discovered factors and association rules of bridge deck deterioration by applying clustering, classification, and association rule data mining algorithms. Although such models showed potential benefits for preventive maintenance, they had limitations in understanding the complex relationship between various damages on bridge elements, such as water leakage to the expansion joint accelerating the corrosion of adjacent girders which lead to weakened bearing [8], and exploring hidden damage mechanism of bridge elements [9, 10].

Sequential pattern is one of useful knowledge that can be derived from data analyses. Sequential pattern mining is a technique to discover patterns of ordered events. It has been widely applied to identify sequential relationship between events and predict posterior events based on the historical patterns in the retail and medicine industries [11, 12]. Since BMS data are periodically recorded and updated, such sequential analyses seem to be applicable to find sequential patterns of damages on bridge elements and estimate future condition changes based on the historical damage patterns. Such sequential patterns can explain damage mechanism between different bridge elements while enhancing prediction performance of structural deterioration of bridges.

Thus, the primary goal of this research is to analyze sequential patterns of damages on the bridge elements. The sequential pattern analysis in the research is summarized as two objectives: (1) determination of structural and environmental characteristics of bridges that cause similar damage patterns, for instance Bridge I and Bridge II follow similar deterioration patterns, and (2) investigation of sequential damage mechanisms such that Damage II has occurred after Damage I. The result of this research will contribute to estimating the future condition changes of the bridges based on the past inspection records for preventive bridge maintenance.

The data used for the analyses was collected from BMS managed by the Korea Institute of Civil Engineering and Building Technology and data mining algorithms for clustering and sequential pattern mining were applied to extract patterns. This paper proposed research methodology for sequential pattern analyses and a preliminary test was performed with a dataset of Pre-Stressed Concrete I-shape type bridges for the validation purpose.

62.2 Research Methodology

To achieve the research goal, the research methodology was organized into three main processes. First, this research collected and preprocessed the data of BMS. Second, the research selected a set of features and performed cluster analysis to discover a number of bridge clusters with distinguished characteristics. Third, sequential pattern mining was utilized to extract types and sequences of damages on bridge elements frequently seen in each cluster. The methodology was developed and implemented by R software version 3.3.2.

62.2.1 Data Collection

The collected BMS data were composed of two structured table datasets: bridge information and bridge inspection records. The bridge information included general, structural, traffic, and weather information of 6773 bridges (i.e., the total of 127 attributes) located in provinces of Korea except Seoul, which were built from 1966 to 2015 (see Table 62.1). Another dataset, the bridge inspection records, contained 834,815 inspection records of the bridge elements. The records were manually entered by inspectors from 1994 to 2015 through 9775 detailed inspections and 900 precise safety diagnoses periodically performed every two to six years. The inspection records included six attributes (i.e., Inspection Bridge Code, Span or Support Code, Inspection Date, Inspection Element, Damage Type, and corresponding Condition Grade). The condition grade was divided into five grades as “A” (best condition), “B”, “C”, “D”, and “E” (worst condition), of which grades “C”, “D”, and “E” meant damaged grades to be repaired (see Table 62.2) [13].

Table 62.1 Sample of bridge information dataset

General information (47 attributes)			Structural information (51 attributes)			Traffic information	Weather information (28 attributes)	
Bridge code	Total length (m)	Total width (m)	Deck type	Deck depth (m)	Girder type	Avg. truck traffic volume (veh./day)	Avg. humidity (%)	Avg. rainy days (/year)
0001	125.2	19.5	RC	20	PSCI	7228	67.7	107
0006	43.8	19.5	RC	60	PSCI	5796	62.4	110
0011	40.8	19.5	RC	23	PSCI	6924	69.2	120

Table 62.2 Sample of bridge inspection records dataset

Inspection bridge code	Span (or support) code	Inspection date	Inspection element	Damage type	Condition grade
0001	01	2006-06-18	Deck	Crack	B
0001	01	2006-06-18	Pavement	Porthole	C
0001	01	2006-06-18	Expansion joint	Corrosion	B
0001	02	2006-06-18	Deck	Crack	B

62.2.2 Data Preprocessing

The research conducted preprocessing before analyses to minimize negative effects of original BMS data on the analyses, which were recorded manually thus included some human errors. Two kinds of preprocessing approaches were applied to a bridge information table. First, the research omitted tuples which contained attributes with missing values to enhance the completeness of dataset, and thus the information of 560 bridges was deleted. Second, the continuous attributes (e.g., traffic volume, length of span, and temperature) were normalized into normal distribution (i.e., mean with zero and standard deviation with one) to avoid the problems caused by different measurement scales on each attribute [14].

The bridge inspection records were also preprocessed by cleaning and reorganizing the attributes. First, the research deleted inspection records with input errors of condition grades such as "F" and "|" instead of "A" to "E". In addition, the records with condition grade of "A" or "B" were also removed since this research only focused on damaged bridge elements: "C", "D", and "E". Second, a new attribute, "Bridge Inspection ID", was made by combining the bridge code and the span or support code. For inspection date, only inspection year was recorded because detailed inspection or safety diagnose is performed only once a year. Consequently, "Inspection year", "Inspection Element", and "Damage Type" were considered for each "Bridge Inspection ID".

62.2.3 Cluster Analysis

Cluster analysis is an unsupervised method to partition a dataset by considering multiple attributes. The key steps include: Feature selection, Clustering algorithm design, and Result interpretation. The research team generated a subset of features from the total 127 attributes of the bridge information table. Then, the clustering algorithm named PAM (Partitioning Around Medoids) was utilized to group bridges with similar features into the same cluster. Finally, the researchers interpreted and discussed clustering results.

PAM introduced by Kaufman and Rousseeuw is one of the most commonly used k-medoids clustering algorithms [15, 16]. The k-medoids algorithm is a clustering approach for partitioning a dataset into k clusters but less sensitive to outliers than a k-means clustering method because each cluster is represented by one of data points of the cluster. These points are named medoids. The k-medoids algorithm needs to specify k, the number of clusters to be generated. In this study, the "pam" function in the "cluster" package on R software was utilized and the optimal k was determined by the silhouette method.

PAM algorithm was proceeded by following steps [16, 17]. Particularly, the study calculated the Gower's dissimilarity matrix which is applied for mixed data types with continuous, ordinal, or categorical attributes at the same time [18].

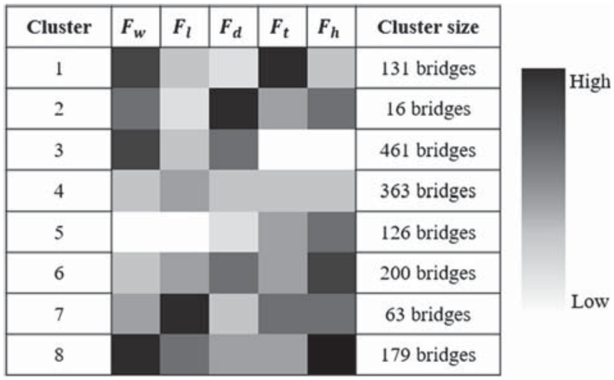

Fig. 62.1 Distinctive characteristics of the features by each cluster

Table 62.3 Example of transaction database based on the bridge inspection records

Bridge inspection ID	Inspection year	Damage types on bridge elements
0001_01	2006	(Damage A, Damage B, Damage C)
0001_01	2010	(Damage B, Damage D)
0001_01	2014	(Damage A, Damage E)
0001_02	2006	(Damage B, Damage C)
0001_02	2010	(Damage A, Damage C, Damage D)

1. Select k objects to become the medoids.
2. Calculate the dissimilarity matrix.
3. Assign every object to its closest medoid.
4. For each cluster, search if any of the object of the cluster decreases the average dissimilarity coefficient.
5. If at least one medoid has changed, go to step 3, else end the algorithm.

62.2.4 Sequential Pattern Mining

Sequential pattern mining (SPM) is a data mining technique to identify patterns of ordered events within transaction database [11]. The bridge inspection records had a large transaction database, where each transaction consisted of three fields: "Bridge Inspection ID" corresponding to the subject of the transaction; "Inspection Year" as transaction time; and "Damage Types on Bridge Elements" as items associated with the transaction (see Table 62.3).

Let $\mathrm{I} = (i_1, i_2, \ldots, i_m)$ be a set of items. A sequence s is an ordered list of item sets. It is denoted by $\mathrm{s} = \langle s_1, s_2, \ldots, s_n \rangle$, where $s_j, j \in 1, 2, \ldots, n$, is an item set, for example $\langle (\mathrm{A}, \mathrm{B}, \mathrm{C}), (\mathrm{B}, \mathrm{D}), (\mathrm{A}, \mathrm{E}) \rangle$. Another sequence $s' = \langle (A), (B), (E) \rangle$, which means Damage B has been produced after Damage A, and subsequently Damage E has occurred, is a subsequence of sequence s since $(\mathrm{A}) \subseteq (\mathrm{A}, \mathrm{B}, \mathrm{C}), (\mathrm{B}) \subseteq (\mathrm{B}, \mathrm{D}), (E) \subseteq (\mathrm{A}, \mathrm{E})$. The support of sequence $s^{'}$ means the proportion of data-sequences which contain $s^{'}$ as subsequence. A minimum support value can be set to decide whether a sequence is frequent or not [6, 19].

SPADE (Sequential Pattern Discovery using Equivalent classes) algorithm introduced by Zaki is one of potential SPM methods. It transforms horizontal transaction database into a vertical id-list database format, which is a list of items consisting of all the IDs and transaction times when the item occurs. This algorithm makes it efficient to reduce database scans in the case of large database such as BMS [20]. The "arulesSequences" package on R software provides an interface to the c++ version of cSPADE.

62.3 Preliminary Test

To validate the proposed research methodology, preliminary test was conducted by using a dataset of 1542 Pre-Stressed Concrete I-shape type (PSCI) bridges with 147,268 inspection records of the superstructure elements. The target elements included expansion joint, deck pavement, deck, girder, and cross beam. After filtering and preprocessing, 31,733 inspection records of the damaged elements were aligned to 3163 PSCI Bridge Inspection ID.

For cluster analysis, five features were selected: effective deck width(F_w), maximum span length(F_l), deck depth(F_d), average truck traffic volume per day(F_t), and average humidity(F_h), which have been known to cause superstructure damages by previous studies [7, 21, 22]. As a result, PSCI bridges were partitioned into eight clusters and the distinctive characteristics of each cluster are illustrated in Fig. 62.1. For example, the bridges in Cluster 1 had high average truck traffic volume per day and thin deck depth compared to the bridges in Cluster 3 which had low average truck traffic volume per day and relatively thick deck depth.

To derive distinct sequential patterns of element damages from each cluster, the authors applied sequential pattern mining algorithms. Different damage types and sequences with 0.03 minimum support were extracted from eight clusters. The minimum support means the ratio of the number of Bridge Inspection IDs that contain such damage types or patterns to the total number of Bridge Inspection IDs within the cluster.

Table 62.4 explains examples of damage types found from Cluster 2 and Cluster 7. The types of severe element damages (e.g., exposed reinforcing steel, breakdown, and deformation) were more frequently discovered in Cluster 2 than Cluster 7. The bridges in Cluster 2 have shorter span length and thicker deck depth than the bridges in Cluster 7, those structural characteristics often explain the bridges are strong against a vertical load but have severe damages instead of common damages such as cracks.

Next, the sequences of damage occurrence were also examined and Table 62.5 shows an example list of the most frequently found sequences within Cluster 7 (long spans and thinner deck depth) in the order by support. Sequence #1 explained same damages can be repeated and similarly Sequence #3 showed the pavement damage could be deteriorated from the crack to the porthole. Sequence #2 examined that the crack of a deck could cause possible leakage from the deck

Table 62.4 Sample damage types on bridge elements seen in Cluster 2 and Cluster 7

Cluster	Damage type on bridge element	Support
2	Exposed reinforcing steel and corrosion of a deck	0.4146
	Concrete efflorescence of a deck	0.2195
	Exfoliation of a girder	0.2195
	Breakdown of a girder anchorage	0.1707
	Deformation of expansion joint	0.0976
7	Crack of a deck	0.2683
	Corrosion of a expansion joint	0.2317
	Exposed reinforcing steel and corrosion of a deck	0.1220
	Crack of pavement	0.1098
	Crack of a cross beam	0.0854

Table 62.5 The most frequently found sequences of Cluster 7

Sequence number	Sequence on bridge elements	Support
1	<(Corrosion of expansion joint), (Corrosion of expansion joint)>	0.2195
2	<(Crack of a deck), (Corrosion of expansion joint)>	0.0854
3	<(Crack of pavement), (Porthole of pavement)>	0.0488
4	<(Rubber-breakdown on expansion joint), (Crack of a cross beam)>	0.0366
5	<(Corrosion of expansion joint), (Exposed reinforcing steel and corrosion of a deck)>	0.0366
6	<(Breakdown of a deck), (Exposed reinforcing steel and corrosion of a deck)>	0.0366

while leading to the second damage of corrosion of expansion joint. Last, Sequence #5 showed that the heavy weight due to long spans could result in weakened expansion joints which then cause severer damages step by step: corrosion of the joint, exposed reinforcing steel, and corrosion of a deck.

62.4 Conclusions

This research aims to analyze sequential patterns of damages on the bridge elements. For the primary goal, the research proposed a research methodology including cluster analysis for discovering bridge clusters with distinguished characteristics and sequential pattern mining for investigating sequential damage causation mechanisms. As preliminary results, PSCI bridges were partitioned into eight clusters and the different sequences of element damages were extracted from each cluster with 0.03 minimum support.

The results showed application potential to estimating future condition changes of the bridges based on the past inspection records for preventive bridge maintenance. However, there are still improvement opportunities. The cluster analysis with more elaborate feature selection needs to be applied to monitor condition changes of the bridges, for instance newly built bridges would follow the deterioration patterns of the aging bridges in the same cluster. In addition, the current research does not explain causal relationships among damage patterns; thus further empirical and statistical analyses need to be conducted to explain damage causation mechanism built on top of the discovered sequential patterns. The further improvement also includes the verification of the pattern-extracting methodology by applying it to other bridges with more various structural types.

Acknowledgements This work was supported by Promising-Pioneering Researcher Program through Seoul National University (SNU) in 2015 and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (Grant No. 2017R1C1B2009237).

References

1. American Society of Civil Engineers (ASCE): Report Card for America's Infrastructure, https://www.infrastructurereportcard.org. last accessed 10 Jan 2018
2. Ministry of Land: Infrastructure, and Transport of Korea: Yearbook of road bridges and tunnels (2016)
3. Agrawal, A.K., Kawaguchi, A., Chen, Z.: Deterioration rates of typical bridge elements in New York. J. Bridg. Eng. **15**(4), 419–429 (2010). https://doi.org/10.1061/(ASCE)BE.1943-5592.0000123
4. Morcous, G.: Performance prediction of bridge deck systems using Markov chains. J. Perform. Constr. Facil. **20**(2), 146–155 (2006). https://doi.org/10.1061/(ASCE)0887-3828
5. Huang, Y.H.: Artificial neural network model of bridge deterioration. J. Perform. Constr. Facil. **24**(6), 597–602 (2010). https://doi.org/10.1061/(ASCE)CF.1943-5509.0000124
6. Kim, Y.J., Queiroz, L.B.: Big Data for condition evaluation of constructed bridges. Eng. Struct. **141**, 217–227 (2017). https://doi.org/10.1016/J.ENGSTRUCT.2017.03.028
7. Huang, R.Y., Chen, P.F.: Analysis of influential factors and association rules for bridge deterioration using national bridge inventory data. J. Mar. Sci. Technol. **20**(3), 336–344 (2012)
8. Sianipar, P., Adams, T.: Fault-tree model of bridge element deterioration due to interaction. J. Infrastruct. Syst. **3**(3), 103–110 (1997). https://doi.org/10.1061/(ASCE)1076-0342
9. Creary, P.A., Fang, F.C.: The data mining approach for analyzing infrastructure operating conditions. Procedia. Soc. Behav. Sci. **96**, 2835–2845 (2013). https://doi.org/10.1016/J.SBSPRO.2013.08.316
10. Liu, K., El-Gohary, N.: Semantic modeling of bridge deterioration knowledge for supporting big bridge data analytics. In: Proceedings of the 2016 Construction Research Congress. pp. 930–939. ASCE Press, San Juan, Puerto Rico (2016)
11. Agrawal, R., Srikant, R.: Mining sequential patterns. In: the 11th International Conference on Data Engineering. pp. 3–14. IEEE Comput. Soc. Press, Taipei, Taiwan (1995)
12. Wright, A.P., Wright, A.T., McCoy, A.B., Sittig, D.F.: The use of sequential pattern mining to predict next prescribed medications. J. Biomed. Inform. **53**, 73–80 (2015). https://doi.org/10.1016/J.JBI.2014.09.003
13. Lim, S., Chung, S., Chi, S., Song, J.: A framework for developing an estimation model of damages on bridge elements using big data analytics. In: the 33rd International Association for Automation and Robotics in Construction. pp. 241–249. I.A.A.R.C., Auburn, AL (2016)
14. Elhag, T.M.S., Wang, Y.M.: Risk assessment for bridge maintenance projects: neural networks versus regression techniques. J. Comput. Civ. Eng. **21**, 402–409 (2007). https://doi.org/10.1016/(asce)0887-3801
15. Kaufman, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis Title. Wiley, United States (1990)
16. Kassambara, A.: Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning. STHDA, United States (2017)
17. Park, H., Jun, C.: A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. **36**(2), 3336–3341 (2009). https://doi.org/10.1016/j.eswa.2008.01.039

18. Podani, J.: Extending gower's general coefficient of similarity to ordinal characters. Int. Assoc. Plant Taxon. (IAPT) **48**(2), 331–340 (1999)
19. Masseglia, F., Poncelet, P., Teisseire, M.: Incremental mining of sequential patterns in large databases. Data Knowl. Eng. **46**(1), 97–121 (2003). https://doi.org/10.1016/S0169-023X(02)00209-4
20. Zaki, M.: SPADE: an efficient algorithm for mining frequent sequences. J. Machine Learning **42**, 31–60 (2001). https://doi.org/10.1023/A:1007652502315
21. Morcous, G., Rivard, H., Hanna, A.M.: Modeling bridge deterioration using case-based reasoning. J. Infrastruct. Syst. **8**(3), 86–95 (2002). https://doi.org/10.1061/(ASCE)1076-0342
22. Huang, R., Mao, I.S., Lee, H.K.: Exploring the deterioration factors of RC bridge decks: a rough set approach. Comput. Civ. Infrastruct. Eng. **25**(7), 517–529 (2010). https://doi.org/10.1111/j.1467-8667.2010.00665.x