A Smart Bridge Data Analytics Framework for Enhanced Bridge Deterioration Prediction

Kaijian Liu[1] and Nora El-Gohary, A.M.ASCE[2]

[1]Assistant Professor, Dept. of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ. E-mail: Kaijian.Liu@stevens.edu
[2]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL. E-mail: gohary@illinois.edu

ABSTRACT

A large amount of data about bridge conditions and maintenance actions, and related factors, are being collected each year. Such data include bridge inventory data, traffic and weather data, and unstructured textual bridge inspection reports. The wealth of these heterogeneous data from multiple sources offers great promise to data analytics for better predicting bridge deterioration. However, existing data-driven bridge deterioration prediction approaches mostly focus on learning from a single type of data from a single source—mainly the National Bridge Inventory (NBI) data. They are limited in learning from multi-type and multi-source data, which collectively cover a large number of factors that affect the deterioration of bridges. To address this limitation, this paper proposes a novel smart bridge data analytics framework. The framework includes three primary components: (1) information extraction: information about bridge conditions and maintenance actions is extracted from unstructured textual inspection reports; (2) data integration: the data/information extracted from the reports are linked and fused, and integrated with bridge inventory, traffic, and weather data; and (3) data analytics: bridge deterioration is predicted based on the integrated data. This paper focuses on presenting the proposed framework and its preliminary experimental evaluation results. The results show that, by learning from integrated bridge data, the proposed framework achieved an average prediction precision and recall of 82.8% and 78.2%, respectively, compared to 71.5% and 60.2% when only learning from NBI data.

INTRODUCTION

Bridges play an important role in ensuring the connectivity of transportation systems for providing daily mobility to the public. However, the U.S. bridges are in critical condition and raise safety concerns. According to the American Society of Civil Engineers (ASCE)'s Infrastructure Report Card, 9.1% and 13.6% of the nation's 614,387 bridges are structurally deficient and functionally obsolete, respectively; and over 188 million daily trips are taken across the deficient bridges (ASCE 2017). It is estimated that the average annual failure rate of the nation's bridges is between 87 and 222, with an expected value of 128 (Cook et al. 2013). Bridge failures, in some cases, are catastrophic. For example, the collapse of the Silver Bridge resulted in 46 fatalities in 1967 (NTSB 1970). The collapse of the I-35W Mississippi River Bridge caused 13 fatalities in 2007 (NTSB 2008). And, the collapse of the Ohio I-75 overpass killed one person in 2015 (Ohio DOT 2015). Such critical conditions of bridges are faced by many countries around the globe, not only by the U.S. alone (Estes 2011). While bridge agencies are striving to improve the conditions of bridges, it is challenging to make cost-effective maintenance decisions under the stringent funding constraints. For example, in the U.S., a $123 billion investment in bridge maintenance is needed to eliminate the nation's deficient bridge backlog, but only $17.5 billion has been invested currently (ASCE 2017). Bridge maintenance decision making relies largely on the predicted future conditions of bridges to allocate the limited maintenance funding in a cost-effective way (Zambon et al. 2017). With such conditions, there has been an increasing demand for data-driven bridge deterioration prediction for supporting enhanced maintenance decision making.

The availability of large amounts of heterogeneous data from multiple sources offers great promise to data analytics for enhanced bridge deterioration prediction for supporting cost-effective maintenance decisions. Such data include structured National Bridge Inventory (NBI) and National Bridge Elements (NBE) data, structured traffic and weather data, and unstructured textual bridge inspection reports. NBI data are bridge-level data, including features about bridge location, geometric characteristics (e.g., bridge length), structural characteristics (e.g., design load), construction characteristics (e.g., year built), and conditions (i.e., the condition ratings of the primary bridge components – decks, superstructures, and substructures). NBE data are element-level data, which describe the condition of a bridge element using four condition states: “good”, “fair”, “poor”, and “severe”. A total of 42 commonly-recognized bridge elements (e.g., deck, top flange, column, wearing surface, etc.) were defined in the NBE data (FHWA 2019a). Traffic and weather data do not directly describe bridge conditions, but are highly relevant to the deterioration of bridges. Traffic data include features about average daily traffic, the percentages of single, double, and triple unit trucks, etc. Weather data include features about cooling and heating degree days, precipitation and snowfall totals, temperature, and diurnal temperature range, etc. Textual inspection reports include a large amount of technically-detailed data/information about bridge conditions and maintenance actions. For example, the following sentence from a report (MnDOT 2006) describes the detailed deficiency conditions (i.e., “minor spalls and patched areas” and “3,000 LF of transverse cracks”) of the bridge element “overlay”: “overlay has some minor spalls and patched areas around the finger joints, and 3,000 LF of transverse cracks”.

To capitalize on the wealth of these data, this paper proposes a novel smart bridge data analytics framework. In the following sections, this paper focuses on presenting the proposed framework and its preliminary experimental evaluation results. The aim of the evaluation is twofold. First, it aims to test the performance of each primary component of the framework. Second, it aims to test if, by learning from the multi-source heterogeneous data, the proposed prediction approach could improve the performance of predicting the future condition ratings of bridge elements – compared to only learning from NBI data.

KNOWLEDGE GAPS IN DATA-DRIVEN BRIDGE DETERIORATION PREDICTION

Despite the availability of a large amount of multi-source heterogeneous data, existing research efforts for data-driven bridge deterioration prediction have mostly focused on utilizing only a limited set of data – mainly abstract bridge inventory data, such as the NBI data which use condition ratings to describe bridge conditions – to predict the deterioration. For example, Morcous et al. (2002) used the bridge inventory data from the Province of Quebec of Canada to develop a case-based reasoning model for predicting the condition ratings of decks. Huang (2011) used the bridge inventory data from the Wisconsin Department of Transportation (DOT) to develop an artificial neural network-based model for predicting the ratings of decks. Zambon et al. (2017) utilized the Markov chain to learn from deck condition ratings from previous years to predict the ratings in the next year. Qiao et al. (2016) developed an ordered binary probit model using the NBI data from the Indiana DOT to predict the condition ratings of the primary bridge components (i.e., decks, superstructures, and substructures). Chang et al. (2018)

developed a stochastic model using the NBI data from the Wyoming DOT to predict the condition ratings of the primary bridge components.

Abstract bridge inventory data are certainly useful, but are not enough; they lack detailed descriptions about bridge conditions and maintenance actions, which limits the ability to learn from history to better predict the future deterioration of bridges. More specifically, existing data-driven prediction methods/models are limited in: (1) making use of the large amounts of rich data/information about bridge conditions and maintenance actions that are buried in textual bridge inspection reports, which misses the opportunity of learning from such rich data for improved deterioration prediction, and (2) utilizing integrated data from multiple sources, which limits the capability to consider a diverse set of factors (e.g., maintenance actions taken, material used in maintenance, traffic and weather patterns, etc.) that affect the deterioration of bridges and are, hence, important to consider when predicting the deterioration.

PROPOSED SMART BRIDGE DATA ANALYTICS FRAMEWORK

To address the aforementioned limitations, the authors propose a novel smart bridge data analytics framework. The proposed framework, as per Figure 1, includes three primary components to allow for the extraction, integration, and analysis of both structured and unstructured bridge data from multiple sources for predicting bridge deterioration: semantic information extraction, data integration, and machine learning-based data analytics.

Semantic Information Extraction

Semantic information extraction aims to extract information that describes bridge conditions and maintenance actions from unstructured textual bridge inspection reports. It is formulated as two tasks: named entity recognition for classifying words in the text into predefined target classes, and dependency parsing for extracting dependency relations from the text to represent the extracted information in a semantically-rich structured way.

An ontology-based, semi-supervised conditional random field algorithm (Liu and El-Gohary 2017) was proposed and used to classify the words in the text into the following target classes: bridge element, deficiency, deficiency cause, maintenance action, maintenance material, numerical measure, numerical measure unit, categorical quantity measure, categorical severity measure, date, and other. The algorithm uses a bridge deterioration knowledge ontology (Liu and El-Gohary 2016) to assist the analysis of the text based on content and domain-specific meaning. The algorithm uses a proposed semantic similarity measure to derive target classes for unlabeled text based on the labeled text, so that the algorithm captures the dependency structures and the distributions of both labeled and unlabeled text and thus adapts itself to unseen text by further learning from a large set of unlabeled text for improved performance of information extraction.

A semantic neural network ensemble algorithm (Liu and El-Gohary 2019) was proposed and used to extract dependency relations from the text for linking the extracted, but still isolated, words into concepts and representing the concepts in a semantically-rich structured way. The algorithm uses the semantics of the text (as captured by the extracted target classes of the words) to capture the word-to-word interactions for facilitating the extraction of dependency relations. The algorithm uses a proposed similarity-based sampling method to sample the configurations of the text that are similarly-distributed and thus more easily-separable into the same cluster. It then uses an ensemble of constituent neural network classifiers – each of which learns from the configurations in the same cluster – to collectively capture the complex distributions of all the configurations and uses a combiner classifier to capture the classification/misclassification patterns of the constituent classifiers to make a final prediction on the dependency relation types.

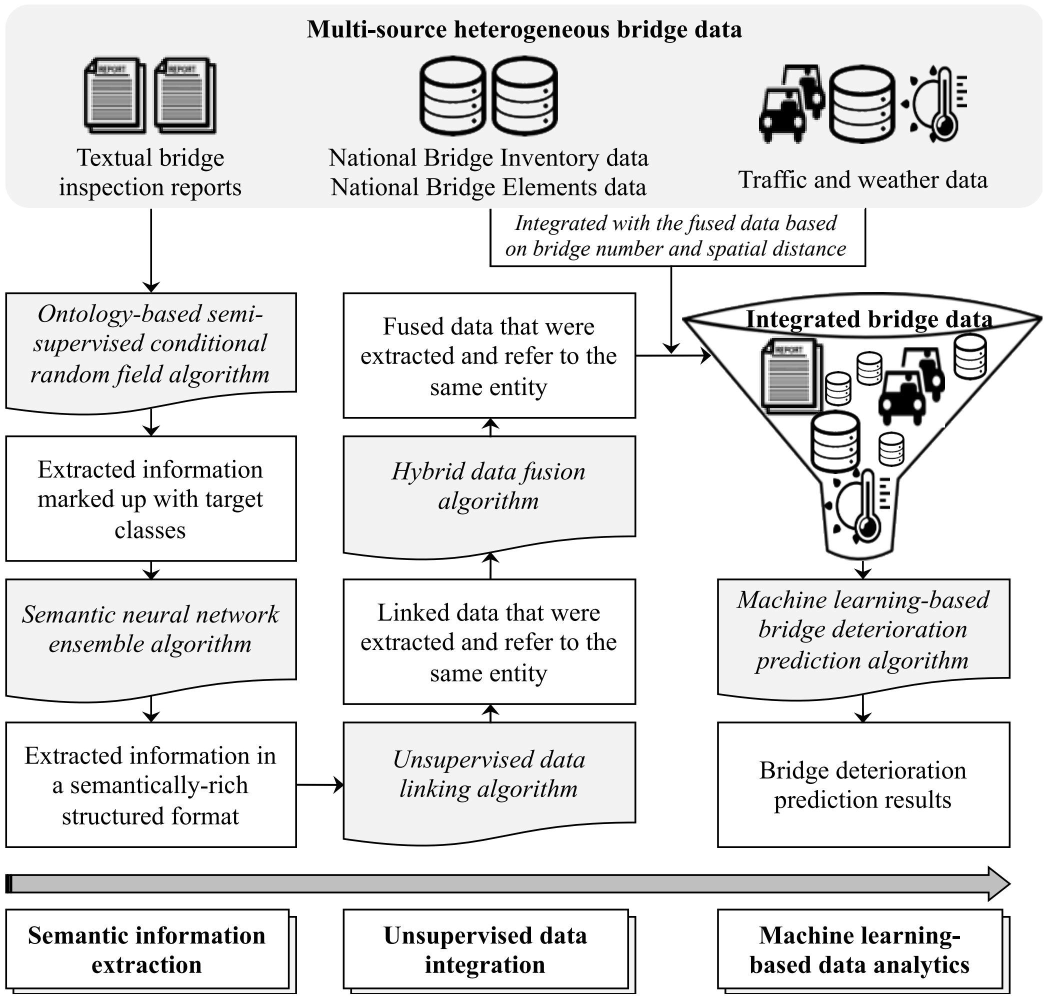

Figure 1. Proposed smart bridge data analytics framework

Unsupervised Data Integration

Unsupervised data integration mainly aims to integrate (i.e., link and fuse) the records that are extracted from textual inspection reports and refer to the same entity (e.g., the same type of deficiency on the same type of bridge element). It is formulated as two tasks: unsupervised data linking for linking the records within each report, and unsupervised data fusion for fusing the linked records from different reports into a unified representation. Data integration, here, also addresses the integration of the fused report data with the other types of structured data, including NBI and NBE data as well as traffic and weather data. The fused report data are integrated with NBI and NBE data based on the bridge structure number, and are integrated with traffic and weather data based on the spatial distances between bridges and traffic/weather monitoring stations. For example, for a bridge, its NBI, NBE, and fused textual data are integrated with the traffic data from the monitoring station that is the closest to the bridge.

An unsupervised data linking algorithm that uses improved similarity assessment and spectral clustering techniques was proposed and used to link the records that refer to the same entity. The improved similarity assessment uses similarity assessment dependencies to break down the record-level similarity assessment task into sequences of attribute-level (i.e., concept-level) tasks. It, then, assesses the similarities between the concepts based on the similarity degrees of their constituent terms, without the need of the contextual information of the concepts from a text corpus or the need of mapping the concepts in comparison to the corresponding concepts in a taxonomy. The improved spectral clustering uses iterative bi-partitioning to automatically identify the optimal number of target clusters (the number of sets containing linked concepts), without using a manually predefined number.

A hybrid data fusion algorithm that combines an unsupervised named entity normalization algorithm and an entropy-based numerical data fusion algorithm was proposed and used to fuse the concept names and the numerical deficiency measures in the linked records. The named entity normalization algorithm aims to fuse the concept names that refer to the same entity but have varied surface forms and abstraction levels into an identifier name that has a canonical form and balances the abstraction and detailedness. It generates candidate canonical names using an n-gram language model, and selects the identifier name from the candidates using corpus statistics and linguistic rules. The numerical data fusion algorithm aims to fuse the numerical measures of multiple deficiency instances of the same type into a single interval-based representation. It defines candidate intervals using the proportional k-interval discretization and fuses the deficiency measures into a candidate interval based on the information entropy of the intervals.

Machine Learning-Based Data Analytics

Machine learning-based data analytics aims to learn from the integrated bridge data – the integrated NBI, NBE, traffic, weather, and textual inspection data – to predict bridge deterioration. This is a challenging machine learning task because the integrated bridge data are highly dimensional and imbalanced. For example, for a single bridge at a timestep in the created dataset (refer to the "Dataset creation" section), its integrated bridge data include a total of 12,687 features. And, bridge data are imbalanced. A large percentage of the bridges are in the condition rating categories of "6" or "7" (i.e., satisfactory and good conditions), and a small percentage of them are in the condition rating categories of "8" (i.e., very good condition), "5", or below "5" (i.e., fair condition or worse). To address these data challenges, a machine learning-based bridge deterioration prediction algorithm was proposed. It uses a manifold learning technique – isometric feature mapping – to embed the high-dimensional bridge data into a low-dimensional and dense space, and uses a focal loss objective function to address the imbalance in the data. The proposed algorithm also employs a feed-forward neural network classifier to learn from the integrated bridge data (embedded and balanced) from past years for predicting the condition ratings of the primary bridge components – decks, superstructures, and substructures – in the next year.

PERFORMANCE EVALUATION OF THE PROPOSED SMART BRIDGE DATA ANALYTICS FRAMEWORK

Dataset Creation

To create a dataset for algorithm training and testing, a total of 1,000 state-owned bridges in the state of Washington were selected. The NBI data were collected from the Federal Highway Administration (FHWA 2019b). The NBE and textual bridge inspection reports were collected from the Washington DOT. The traffic data were collected from the Transportation Data, GIS and Modeling Office of the Washington DOT (WSDOT 2019). The weather data were collected from the National Oceanic and Atmospheric Administration (NOAA 2019).

Performance Evaluation

To evaluate the performances of the information extraction and data integration, a gold standard for each algorithm/model of the two components was, separately, prepared by three annotators. A subset of the bridge inspection reports were randomly selected for annotation due to the infeasibility of manually annotating all the reports in the dataset. The final gold standard for each algorithm/model (achieved with full agreement of all the annotators) was compared to the algorithm-generated results based on precision and recall. To evaluate the performance of the data analytics, the machine learning-based prediction algorithm was trained and tested using 3-fold cross validation. In the testing fold, the algorithm-predicted condition ratings were compared to the gold standard ratings (i.e., the real condition ratings as reported in the NBI data) based on precision and recall. Precision and recall are calculated using Eqs. (1) and (2), where TP, FP, and FN are the number of true positives, false positives, and false negatives, respectively. Depending on the component of the proposed framework to be tested, TP and FP could be the number of correctly and incorrectly extracted information entities, linked records, or predicted condition ratings, respectively. FN could be the number of entities, records, or ratings that should be, but were not, extracted/linked/predicted by the respective algorithms.

$$Precision = \frac{TP}{(TP + FP)} \tag{1}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{2}$$

Performance Results

The performance results of the components of the proposed smart bridge data analytics framework are summarized in Table 1. Overall, the proposed framework performed well: it achieved an average precision and recall of 82.8% and 78.2%, respectively, when predicting the condition ratings of the primary bridge components. Two main sources of errors that contributed to the incorrectly-predicted condition ratings were identified. First, the errors occurred at the information extraction and data integration steps could have propagated into the data analytics step, which negatively affected the prediction performance. The information extraction component achieved a precision and recall of 88.7% and 85.5%, respectively. Incorrect and missed extractions negatively affected the performance of the following data integration component (i.e., precision = 95.9% and recall = 83.2%), which subsequently introduced data errors and noise to the analytics step. Second, the machine learning-based prediction algorithm was not able to fully deal with the high dimensionality and the imbalance in the integrated bridge data. Although it uses manifold learning and focal loss objective function to address these data challenges to a great extent, like all other machine learning algorithms, the prediction algorithm cannot fully deal with the negative impacts caused by these challenges. As a result, when fitting hyperplanes in a somewhat still high-dimensional and imbalanced space to separate different condition rating patterns, the algorithm "inevitably" generated some prediction errors.

Comparison With Existing Approaches

Existing bridge deterioration prediction approaches mostly learn from NBI data (or similar bridge inventory data collected by different countries) to predict the condition ratings of decks, superstructures, and substructures. The two types of bridge deterioration prediction approaches (i.e., the proposed and existing) were compared to test if learning from integrated multi-source heterogeneous bridge data, compared to learning from NBI data alone, could improve the prediction performance. Table 2 summarizes the comparison results.

The comparison results show that learning from the integrated multi-source heterogeneous bridge data outperformed learning from the NBI data alone. On average, it achieved a precision and recall of 82.8% and 78.2%, respectively, compared to 71.5% and 60.2% achieved by only learning from the NBI data. NBI data mainly include features about the as-built characteristics of bridges (e.g., geometric, structural, and construction characteristics) and describe bridge conditions using condition ratings. Although NBI data are very important, they are limited in capturing the deterioration patterns of bridges, because they lack detailed descriptions about the deficiency conditions and maintenance actions of bridges. Depending on different types of conditions and maintenance, bridges with same/similar as-built characteristics could be in different condition rating categories. In such case, learning from the integrated bridge data, especially previously-untapped textual inspection reports which include a wealth of detailed data/information about bridge conditions and maintenance actions, was able to better capture the deterioration patterns of the bridges and, thus, achieved improved performance in predicting the condition ratings.

Table 1. Performance Results of the Proposed Smart Bridge Data Analytics Framework

Framework component	Algorithm	Precision (%)	Recall (%)
Semantic information extraction	Ontology-based semi-supervised conditional random field algorithm	95.7	87.3
	Semantic neural network ensemble algorithm	88.7	85.5
Unsupervised data integration	Unsupervised data linking algorithm	87.8	84.9
	Unsupervised named entity normalization algorithm	95.9	83.2
	Entropy-based numerical data fusion algorithm	–[1]	–[1]
Machine learning-based data analytics	Machine learning-based bridge deterioration prediction algorithm	82.8	78.2

[1] Not applicable since there is no gold standard for numerical data fusion.

Table 2. Performance Results of the Proposed Bridge Deterioration Prediction Approach in Comparison with Existing Approaches

Approach	Performance in predicting the condition ratings for primary bridge components					
	Deck		Superstructure		Substructure	
	Precision	Recall	Precision	Recall	Precision	Recall
Proposed [1]	82.4%	76.3%	84.4%	78.5%	81.7%	79.8%
Existing [2]	72.0%	62.5%	72.2%	59.1%	70.2%	59.1%

[1]The proposed bridge deterioration prediction approach/framework, i.e., learning from the integrated multi-source heterogeneous bridge data.
[2]Existing approaches, i.e., only learning from the National Bridge Inventory (NBI) data.

CONCLUSION AND FUTURE WORK

In this paper, the authors proposed a novel smart bridge data analytics framework to allow for the extraction, integration, and analysis of both structured and unstructured bridge data from multiple sources for enhanced bridge deterioration prediction. Such data include structured NBI and NBE data, structured traffic and weather data, and unstructured textual bridge inspection reports. The proposed framework includes three primary components: information extraction, data integration, and machine learning-based data analytics. The performance of the proposed framework was evaluated. The experimental results show that the proposed framework performed well: it achieved an average precision and recall of 82.8% and 78.2%, respectively, when predicting the condition ratings of decks, superstructures, and substructures. The results also show that, by learning from the integrated multi-source heterogeneous bridge data, the proposed approach improved the prediction precision and recall by 11.4% and 18.0%, respectively, compared to only learning from NBI data. In their ongoing/future research, the authors will further test the performance of the proposed framework in predicting the quantities of specific bridge element-level deficiencies, which would allow for better allocating the resources to the right bridge deficiencies that require imminent maintenance. Along this line of research, the authors will also explore the use of bridge inspection images and health monitoring data, in integration with the multi-source heterogeneous data used in this study, for better supporting bridge deterioration prediction.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Foundation (NSF). This paper is based upon work supported by NSF under Grant No. 1937115. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF. Funding for this research was also provided in part by the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign through the NCSA Faculty Fellows program. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF's Grant No. ACI-1548562.

REFERENCES

ASCE (American Society of Civil Engineers). (2017). "2017 Infrastructure report card." <http://www.infrastructurereportcard.org/>. (Jul. 30, 2019).

Chang, M., Maguire, M., and Sun, Y. (2018). "Stochastic modeling of bridge deterioration using classification tree and logistic regression." J. Infra. Sys., 25(1).

Cook, W., Barr, P. J., and Halling, M. W. (2013). "Bridge failure rate analysis." Proc., Transp. Res. Board 92nd Annual Meeting, Washington D.C., 13-27.

Estes, A. C. (2011). "Bridge maintenance, safety, management, and life-cycle optimization." J. Struc. Infra. Engr., 7(10):807.

FHWA (Federal Highway Administration). (2019a). "Specification for the national bridge inventory bridge elements." <https://www.fhwa.dot.gov/bridge/nbi.cfm> (Jul. 30, 2019).

FHWA (Federal Highway Administration). (2019b). "National bridge inventory." <https://www.fhwa.dot.gov/bridge/nbi/ascii.cfm> (June 21, 2019).

Huang, Y. H. (2010). "Artificial neural network model of bridge deterioration." J. Perf. Constr. Facil., 24(6):597-602.

Liu, K., and El-Gohary, N. (2016). "Semantic modeling of bridge deterioration knowledge for supporting big bridge data analytics." Proc., Constr. Res. Congr., ASCE, Reston, VA., 930-939.
Liu, K., and El-Gohary, N. (2017). "Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports." Autom. Constr., 81:313-327.
Liu, K., and El-Gohary, N. (2019). "Semantic neural network ensemble for automated dependency relation extraction from bridge inspection reports." Autom. Constr., submitted.
MnDOT (Minnesota Department of Transportation). (2006). "Fracture critical bridge inspection in-depth report: I-35W over the Mississippi River at Minneapolis, Minnesota." Saint Paul, MN.
Morcous, G., Rivard, H., and Hanna, A. M. (2002). "Modeling bridge deterioration using case-based reasoning." J. Infra. Sys., 8(3):86-95.
NOAA (National Oceanic and Atmospheric Administration). (2019). "Climate data online." <https://www.ncdc.noaa.gov/cdo-web/> (Sep. 14, 2019).
NTSB (National Transportation Safety Board). (1970). "Collapse of U.S. Highway Bridge, Point Pleasant, West Virginia." <https://www.ntsb.gov/investigations/AccidentReports/Pages/HAR7101.aspx> (Jan. 7, 2019).
NTSB (National Transportation Safety Board). (2008). "Highway accident report Interstate 35W over the Mississippi River Minneapolis, Minnesota." <https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR0803.pdf> (Jan. 7, 2019).
Ohio DOT (Ohio Department of Transportation). (2015). "Background/timeline information on the Interstate 75/Hopple Street Ramp collapse." <http://www.dot.state.oh.us/news/i-75-cincinnati-bridge-information/Pages/default.aspx> (Jan. 7, 2019).
Qiao, Y., Moomen, M., Zhang, Z., Agbelie, B., Labi, S., and Sinha, K. C. (2016). "Modeling deterioration of bridge components with binary probit techniques with random effects." Transp. Res. Rcrd., 2550(1):96-105.
WSDOT (Washington Department of Transportation). (2019) "Traffic data GeoPortal." <https://www.wsdot.wa.gov/mapsdata/tools/trafficplanningtrends.htm> (Sep. 14, 2019).
Zambon, I., Vidovic, A., Strauss, A., Matos, J., and Amado, J. (2017). "Comparison of stochastic prediction models based on visual inspections of bridge decks." J. Civil Engr. Mgmt., 23(5):553-561.