Learning from Class-Imbalanced Bridge and Weather Data for Supporting Bridge Deterioration Prediction

90

Kaijian Liu and Nora El-Gohary

Abstract

Evaluating the impact of learning from weather data, in addition to bridge data, on the performance of bridge deterioration prediction is critical for identifying the right data needed for better prediction for enhanced bridge maintenance decision making. However, the majority of the studies in the bridge domain did not consider such evaluation. For those that conducted the evaluation, their evaluation results usually varied. There is, thus, a need for re-evaluating whether the use of weather data could improve the prediction performance. However, conducting the evaluation is challenging because of class imbalance problems in the bridge domain. Therefore, prior to the evaluation, conducting data sampling to alleviate/eliminate such problems is necessary. To address these needs, this paper offers a pilot evaluation study for better evaluating the impact of learning from weather data on bridge deterioration prediction. To conduct the evaluation, a sampling method was used to deal with the data imbalance problems, and a deep neural network model was developed to predict the condition ratings of decks, superstructures, and substructures. A number of alternative sampling methods were tested and the prediction performances—with and without weather data—were compared. The preliminary experimental results indicated that: (1) the random over-sampling method outperformed the other alternatives; and (2) the change in the prediction performance after further learning from the weather data was only marginal.

Keywords

Weather and bridge data • Bridge deterioration prediction • Data imbalance problems • Deep neural networks

90.1 Introduction

Bridge deterioration prediction is an indispensable component of bridge maintenance decision making. Recent studies (e.g., [1–3]) have emphasized the need for taking a data-driven approach for better predicting bridge deterioration. The increasing availability of heterogeneous bridge data from different sources opens unprecedented opportunities to data analytics for better predicting deterioration and for learning how to better maintain bridges. Such data include National Bridge Inventory (NBI) and National Bridge Elements (NBE) data, traffic and weather data, unstructured sensory data from various types of sensors, and unstructured textual data from bridge inspection reports.

To capitalize on the wealth of these data, in their previous work [4–6], the authors proposed a big bridge data analytics framework. It includes three main components: (1) semantic information and relation extraction for extracting information about bridge conditions and maintenance actions from unstructured textual bridge inspection reports, and representing the extracted information in a structured way [4], (2) semantic data linking and fusion for integrating data from multiple sources into a unified representation [5], and (3) semantic data analytics for predicting bridge deterioration and learning bridge

K. Liu (✉) · N. El-Gohary
Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: kliu15@illinois.edu

N. El-Gohary
e-mail: gohary@illinois.edu

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_90

maintenance strategies [6]. One of the most important research tasks in the data analytics component is to evaluate the impact of learning from weather data on the performance of data-driven, machine learning (ML)-based bridge deterioration prediction. This is because of two main reasons. First, such an evaluation can help identify the right data needed for better predicting bridge deterioration. Second, as further analyzed in Sect. 2.2, the impacts of learning from weather data concluded by previous studies varied or even disagreed.

There is, thus, a need for re-evaluating whether learning from weather data, in addition to bridge data, could improve the prediction performance. However, conducting such an evaluation is challenging because of the class imbalance problems. The numbers of bridges in different condition rating categories are naturally imbalanced. Such imbalance negatively affects the prediction performance. Further incorporating weather data as features for evaluating its impact on bridge deterioration prediction increases the data dimensionality. This changes the levels of the impacts caused by the imbalance [7] and thus makes the prediction performances across the datasets incomparable. Previous studies are rather limited in considering the varied impacts of the imbalance, which could be the key contributing factor that led to the conflicting evaluation results. Therefore, prior to the evaluation, conducting data sampling to alleviate/eliminate such varied impacts across the datasets is necessary. To address these needs, this paper offers a pilot evaluation study for better evaluating the impact of learning from weather data—in addition to bridge data—on the performance of bridge deterioration prediction. This paper focuses on presenting the evaluation method and the preliminary experimental results.

90.2 Background

90.2.1 State of the Art in Bridge Deterioration Prediction

Many research efforts have been undertaken towards developing data-driven bridge deterioration prediction methods/models. The majority of them focused on using bridge characteristic data for developing bridge deterioration prediction models, and did not consider weather data or sufficiently evaluate their impacts on the prediction. For example, Huang [8] developed an artificial neural network model for predicting the condition ratings of bridge decks. The model mainly learns from data about bridge characteristics, such as design load, deck length, and number of spans. Li et al. [9] developed a Markov-Chain-based bridge deterioration model to capture the transition probabilities of bridge condition ratings for predicting urban bridge deterioration at the network level.

On the other hand, there are several studies (e.g., [10, 11]) that evaluated the impact of using weather data on the performance of bridge deterioration prediction. However, their evaluation results varied largely. For example, Chang et al. [10] concluded that weather data were found not important. However, Qiao et al. [11] concluded that the region variable (a proxy for climate/weather) is generally influential, and the number of cold days is among the most significant factors. In addition to these studies, there are studies (e.g., [12, 13]) that tried to identify the critical sources/factors of bridge deterioration. For example, Kim and Yoon [12] identified the following factors as the most significant contributors to bridge deterioration: year built, structural characteristics, and traffic volume. Weather data, such as precipitation, snow fall, and average temperature, were not among these most-contributing factors [12]. But, Huang et al. [13] found that factors related to weather were rather significant in bridge deterioration.

90.2.2 Knowledge Gaps

Despite the importance of the existing research efforts, two main knowledge gaps are identified. First, the majority of the existing studies did not evaluate the impact of using weather data on the performance of data-driven bridge deterioration prediction. Such evaluation is, however, very important for identifying the right source of data to be used in data-driven bridge deterioration prediction. Using the right data without abundancy is key for improving the prediction performance and reducing the computational needs. Second, although there are several studies that attempted to evaluate the impact of using weather data on the performance of data-driven bridge deterioration prediction, they are rather limited in considering the varied impacts across datasets caused by the imbalance. As discussed, the numbers of bridges in different condition rating categories are naturally imbalanced. Further incorporating weather data as features for evaluating its impact increases the data dimensionality. This changes the levels of the negative impacts caused by the imbalance and thus makes the prediction performances across the datasets incomparable. The existing studies, mostly, did not consider the varied impacts of the imbalance, which could be the key contributing factor that led to the conflicting evaluation results.

90.3 Evaluation Method

To address the above-mentioned knowledge gaps, this paper offers a pilot evaluation study for better evaluating the impact of learning from weather data, in addition to bridge data, on the performance of data-driven, ML-based bridge deterioration prediction. To conduct the evaluation, a sampling method was used to deal with the data imbalance problems, and a deep neural network (DNN) model was developed to predict the condition ratings of decks, superstructures, and substructures. The prediction performances—with and without weather data—were evaluated. The evaluation study, thus, included five main steps: (1) data collection, (2) data preprocessing, (3) data sampling, (4) DNN modeling, and (5) performance evaluation.

90.3.1 Data Collection

Two main types of data were collected for this pilot study: data about bridge characteristics and data about weather characteristics. The bridge data included features about bridge location, geometric characteristics (e.g., bridge length, deck width, number of spans, etc.), structural characteristics (e.g., functional classification, design load, wearing surface type, etc.), construction characteristics (e.g., year built and type of construction), and traffic volumes (e.g., average daily traffic and percent of truck traffic). These data were collected from the National Bridge Inventory of the Federal Highway Administration. The weather data included features about cooling degree days, heating degree days, diurnal temperature range, precipitation totals, snowfall totals, and temperature. Table 90.1 provides a brief description for the main weather features. These data were collected from the National Oceanic and Atmospheric Administration (NOAA). As a pilot study, this work focused on the bridge and weather data collected from nine U.S. states: Illinois (IL), Michigan (MI), Delaware (DE), Idaho (ID), Arkansas (AR), South Carolina (SC), Utah (UT), Nevada (NV), and Wyoming (WY). These states were selected because each of them belongs to one of the climatically consistent regions that were defined by the NOAA. By capturing different weather and bridge deterioration patterns, the collected data would, thus, help better evaluate the impact of learning from weather data.

As a result, a total of 1078 data instances were included in this study. Three different datasets were then generated. Dataset #1 only included the data about bridge characteristics. Dataset #2 included the bridge data as well as partial weather data (i.e., only the weather features about the normal were included). For example, for the cooling degree days, only the feature about the cooling degree day normal, which is computed with a base of 65 °F, was included. Dataset #3 included both the bridge data and the full weather data (i.e., all the weather features were included).

Table 90.1 A brief description of the main features of the weather data

Feature	Brief description
Cooling degree days	The cooling degree day normal is computed with a base of 65 °F. The other bases include 45, 50, 55, 57, 60, 70, and 72 °F
Diurnal temperature range	The diurnal temperature range is the difference between the F maximum and minimum temperature over a day
Heating degree days	The heating degree day normal is computed with a base of 65 °F. The other bases follow those used in cooling degree days
Precipitation totals	The annual precipitation totals. The other features about the precipitation include the number of days whose precipitation is greater than 0.01, 0.10, 0.50, and 1.00 inches
Snowfall totals	The annual snowfall totals. Similarly, the other features include the number of days whose snowfall is greater than 0.1, 1.0, 3.0, 5.0, and 10.0 inches
Average temperature	The annual average temperature
Maximum temperature	The annual maximum temperature. The other features include the number of days whose maximum temperature is greater than 32, 40, 50, 60, 70, 80, 90, and 100 °F
Minimum temperature	The annual minimum temperature. Similarly, the other features include the number of days whose minimum temperature is less than 32, 40, 50, 60, and 70 °F

90.3.2 Data Preprocessing

Four main steps were conducted to process the weather and bridge data. First, the missing data values in the data were imputed. For each feature type, the mean of the data values that are available was calculated and then used for imputing the missing ones. Second, the data values of the numerical features (e.g., bridge length and average temperature) were normalized. Both the original and imputed values of the same feature type were translated into the same range between 0 and 1, using the min-max normalization method. Third, the data values of the categorical features (e.g., wearing surface type) were converted into numerical values, using one-hot feature representation. Finally, the bridge data were mapped to the weather data. In conducting the mappings, for each weather station, its distances to all the bridges were calculated. The station was mapped to the bridge that is spatially closest. For calculating the distance, the latitudes and longitudes of the bridges and the weather stations were used.

90.3.3 Data Sampling

Data sampling aimed to balance the training dataset for addressing the negative, yet varied, impacts caused by the imbalance. Table 90.2 summarizes the imbalance characteristics of the data. The numbers of bridges in different condition rating categories (the ratings are the target classes for the bridge deterioration prediction) are naturally imbalanced. For example, there are around 30.5 and 8.0% of the bridges whose deck condition ratings are 7 and 5, respectively. Two main sampling approaches are commonly used for addressing data imbalance problems: over-sampling and under-sampling. The over-sampling approach [e.g., the random over-sampling technique and the synthetic minority over-sampling technique (SMOTE)] adds data instances to the minority classes to create a balanced dataset [14]. The under-sampling approach (e.g., the random under-sampling and the cluster centroid-based techniques) removes a subset of the data from the majority classes for balancing the dataset [14]. For this pilot study, several commonly-used sampling methods were tested and compared. These methods were selected based on the state-of-the-art literature review studies [14, 15] in the area of imbalanced learning. Table 90.3 provides a summary of the selected sampling methods/techniques.

90.3.4 Deep Neural Network Modeling

A deep neural network (DNN) model was developed for learning from the sampled data instances for predicting the condition ratings of the decks, superstructures, and substructures. This deep learning approach was selected mainly because it is the current state-of-the-art ML approach and has been applied for bridge deterioration prediction (e.g., [8]) and many other ML applications (e.g., [16]). In this paper, the DNN model was developed with a feed-forward architecture that

Table 90.2 Imbalance characteristics of the dataset

Condition rating[a]	Percentage of bridge elements in different condition rating categories		
	Deck (%)	Superstructure (%)	Substructure (%)
N	22.6	21.6	21.6
0	0.2	0.1	0.1
1	0.0	0.0	0.0
2	0.1	0.1	0.1
3	0.6	0.9	0.6
4	1.9	2.0	2.2
5	8.0	9.2	8.4
6	24.5	21.1	22.2
7	30.5	27.8	31.2
8	10.2	15.4	11.8
9	1.6	1.8	1.8

[a]Condition rating of "0" stands for "failed condition"; condition rating of "9" stands for "excellent condition"; condition rating of "N" stands for "not applicable"

Table 90.3 Commonly-used candidate sampling methods

No.	Method	Description[a]
SM #1	Random over-sampling	Randomly select samples from the minority classes, and augment the dataset by replicating the selected samples and adding the replicates to the dataset
SM #2	Synthetic minority over-sampling technique (SMOTE) 1	For a sample from the minority classes, its nearest neighbors are defined based on the Euclidian distance. The sample that has half of its neighbors belong to a different class will be used for the sampling. A new synthetic sample will be generated by interpolating between the original sample and one of its neighbors who has a different class
SM #3	SMOTE 2	Similar to SMOTE 1. But, a new synthetic sample will be generated by interpolating between the original sample and any one of its neighbors
SM #4	Random under-sampling	Randomly select samples from the majority classes, and remove the samples from the dataset
SM #5	Cluster centroids	A number of clusters are generated based on the samples in each of the majority classes. The number is defined based on the smallest number of samples in the minority cluster. The centroids of the clusters are used as samples for representing the majority classes
SM #6	Near miss	Select the samples from the majority classes, and use a subset of the samples whose average distance to the furthest samples of the minority classes is the smallest

[a]SM = sampling method. The description is based on literature review studies [14] and [15]

contains an input layer, several hidden layers, and a softmax output layer. Different numbers of hidden layers and activation functions were tested through model parameter tuning. As a result, the DNN model with five hidden layers and the rectified linear unit activation function was developed.

90.3.5 Performance Evaluation

Two performance evaluation cases were developed. The first case aimed to compare the performances of the selected sampling methods. The second one aimed to evaluate the impact of learning from the weather data, in addition to the bridge data, on the performance of bridge deterioration prediction—predicting the condition ratings of the decks, superstructures, and substructures. Accuracy was selected as the evaluation metric for benchmarking the prediction performance. Accuracy, here, is the percentage of the number of correctly-predicted condition ratings out of the total number of the predicted ratings. In order to examine the robustness of the results, for each experimental run, a 10-fold cross validation was conducted. For each validation iteration, the data instances in the training folds were sampled by the tested sampling method, and the resulting dataset was used for training a DNN model. The performance of the sampling method and the impact of learning from the weather data were tested using the instances in the testing fold. The average accuracy over the ten folds was reported.

90.4 Preliminary Experimental Results and Discussion

90.4.1 Performances of Data Sampling Methods

As mentioned in Sect. 3.2, a total of six commonly-used data sampling methods for dealing with the class imbalance problems were tested. Table 90.4 summarizes their performances for the three datasets. Three main observations are drawn from the experimental results. First, the over-sampling approach (SMs #1, #2, and #3 as per Table 90.4) performed better than the under-sampling approach (SMs #4, #5, and #6). On average, compared to the baseline where no sampling method was used, the over-sampling approach improved the accuracy by over 30.0%. The under-sampling approach only improved the accuracy by around 14.0%. This is mainly because the under-sampling methods decreased the size of the dataset. For

Table 90.4 The performances of the data sampling methods

Method[a]	Dataset #1 (%)	Dataset #2 (%)	Dataset #3 (%)	Average (%)
Baseline	52.1	51.6	53.2	52.3
Baseline + SM #1	+36.9[b]	+37.7	+34.7	+36.4
Baseline + SM #2	+32.8	+33.0	+30.9	+32.2
Baseline + SM #3	+30.6	+31.2	+29.0	+30.3
Baseline + SM #4	+14.5	+14.3	+14.3	+14.4
Baseline + SM #5	+14.1	+14.7	+13.9	+14.3
Baseline + SM #6	+13.2	+15.6	+14.6	+14.5

[a]The baseline was developed without using any sampling method. The index for the sampling methods follows that defined as per Table 90.3 (SM = sampling method)
[b]The average prediction accuracy improvement compared to the baseline

example, the random under-sampling method randomly dumped the data instances of the majority classes to make the numbers of the instances in both majority and minority classes equal. The decrease in the data size prevented the DNN model to sufficiently capture the different deterioration patterns represented by the dumped instances. Conversely, the over-sampling approach boosted the dataset by adding more instances to the minority classes, which preserved all the patterns represented by the data, and, at the same time, addressed the imbalance problems.

Second, the random over-sampling method outperformed the other over-sampling methods. On average, it improved the prediction accuracy by 36.4%, which is 4.2 and 6.1% higher compared to SMs #2 and #3, respectively. This indicates that: (1) sampling the data by interpolation, as used in SMs #2 and #3, created synthetic data that cannot fully represent the real-world deterioration cases/patterns; and (2) the interpolation generated some data instances that are close to the decision boundaries for separating different condition ratings. This made the classifiers somewhat confused and thus resulted in less-improved performance. Third, although the under-sampling approach reduced the data size, it still performed better than the baseline. This shows that, compared to the reduced data size, the imbalance is more significant in negatively affecting the performance of the bridge deterioration prediction.

90.4.2 Impact of Weather Data on Bridge Deterioration Prediction

The impact of learning from the weather data, in addition to the bridge data, on the performance of the bridge deterioration prediction was evaluated. The evaluation was conducted by comparing the prediction accuracies of the three datasets. The evaluation results are summarized in Table 90.5.

The evaluation results indicate that the impact of using the weather data on the prediction performance is marginal. Using the bridge data alone (i.e., dataset #1) an average prediction accuracy of 89.1% was achieved. Compared to this accuracy, adding the partial weather data (i.e., dataset #2) only improved the accuracy by 0.2% (i.e., an accuracy of 89.3%). Further adding all the weather data (i.e., dataset #3), however, decreased the accuracy by 1.3%. As seen, the increase and the decrease rates are all around 1.0%. This indicates that the impact of learning from the weather data—in addition to the bridge data—on the performance of the prediction, whether positive or negative, is marginal. This could be largely attributed to the low discriminating-power of the weather data. The weather data in a given region are quite similar because of geographical

Table 90.5 The impact of learning from weather data on the performance of the data-driven, ML-based bridge deterioration prediction

Bridge element	Bridge deterioration prediction accuracy		
	Bridge data only (%)	Bridge data + partial weather data (%)	Bridge data + full weather data (%)
Deck	89.4	89.7	88.7
Superstructure	87.9	88.2	86.6
Substructure	90.0	90.1	88.3
Average[a]	89.1	89.3	87.8

[a]The average prediction accuracy over the bridge elements

proximity. But, the bridges in the same region do have different condition ratings. Therefore, the weather data did not improve the ability to distinguish different deterioration patterns—on the contrary, the data even sometimes introduced noise to the classifiers (e.g., when the dimension of the additional weather features is high). As a result, learning from weather data only marginally impacted the performance of the bridge deterioration prediction.

90.5 Conclusions, Limitations, and Future Work

A pilot evaluation study was conducted to better evaluate the impact of learning from weather data, in addition to bridge data, on the performance of data-driven, ML-based bridge deterioration prediction. A set of experiments were conducted to select a suitable sampling method that can deal with the data imbalance, and to compare the prediction accuracies, with and without weather data. The experimental results show that the random over-sampling method, although simple, was effective and suitable for dealing with the imbalance. The results also indicate that the impact of further learning from the weather data, in addition to the bridge data, could be positive or negative, depending on the types of weather data used. However, in either case, the impact was found marginal.

One main limitation of this paper is acknowledged. For this pilot study, the bridge data only included those about the bridge characteristics collected from 2017 NBI. When training/testing the prediction model, the dataset was split using a 10-fold cross validation. As a result, the applicability of the model was limited to learning from static bridge characteristic data to predict the condition ratings of decks, superstructures, and substructures that were not inspected during an inspection year. While this model is sufficient for the purpose of conducting the pilot evaluation study, it would not be sufficient to adequately support bridge maintenance decision making, especially when it comes to providing information about which defects that a bridge could develop in the future and which maintenance action is most cost-effective to repair existing defects and prevent the predicted potential defects.

In their ongoing/future work, the authors will focus on developing new ML algorithms for learning from heterogeneous bridge data from different sources, in order to better predict bridge deterioration for enhanced bridge maintenance decision making. Two main directions will be explored. First, the authors will explore ways to use ML to capture bridge deterioration patterns across time for predicting bridge deterioration—in terms of condition ratings, defect types, and maintenance action types—in the future years. Second, while learning from a big size of the bridge data, the authors will further test the performances of sampling methods in supporting such a complex ML-based prediction task.

Acknowledgements This material is based upon work supported by the Strategic Research Initiatives Program by the College of Engineering at the University of Illinois at Urbana-Champaign.

References

1. Rafiq, M.I., Chryssanthopoulos, M.K., Sathananthan, S.: Bridge condition modelling and prediction using dynamic Bayesian belief networks. Struct. Infrastruct. Eng. **11**(1), 38–50 (2015)
2. Morcous, G., Lounis, Z., Cho, Y.: An integrated system for bridge management using probabilistic and mechanistic deterioration models: application to bridge decks. KSCE J. Civ. Eng. **14**(4), 527–537 (2010)
3. Federal Highway Administration: The LTBP deterioration modeling algorithm. https://www.fhwa.dot.gov/publications/ltbpnews/15073.cfm. Accessed 10 Apr 2018
4. Liu, K., El-Gohary, N.: Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports. Autom. Constr. **81**, 313–327 (2017)
5. Liu, K., El-Gohary, N.: Unsupervised named entity normalization for supporting information fusion for big bridge data analytics. In: Smith, I., Domer, B. (eds.) Proceedings of 25th International Workshop on Intelligent Computing in Engineering, pp. 130–149. Springer, Cham (2018)
6. Liu, K., El-Gohary, N.: Feature discretization and selection methods for supporting bridge deterioration prediction. In: Proceedings of the 2018 ASCE Construction Research Congress, pp. 413–423. American Society of Civil Engineers, Reston (2018)
7. Lusa, L.: Class prediction for high-dimensional class-imbalanced data. BMC Bioinform. **11**, 523 (2010)
8. Huang, Y.: Artificial neural network model of bridge deterioration. J. Perform. Constructed Facil. **24**(6), 597–602 (2010)
9. Li, L., Sun, L., Ning, G.: Deterioration prediction of urban bridges on network level using Markov-chain model. Math. Probl. Eng. (2014)
10. Chang, M., Maguire, M., Sun, Y.: Framework for mitigating human bias in selection of explanatory variables for bridge deterioration modeling. J. Infrastruct. Syst. **23**(3), 04017002 (2017)
11. Qiao, Y., Moomen, M., Zhang, Z., Agbelie, B., Labi, S., Sinha, K.C.: Modeling deterioration of bridge components with binary probit techniques with random effects. J. Transp. Res. Board, 96–105 (2016)
12. Kim, Y.J., Yoon, D.K.: Identifying critical sources of bridge deterioration in cold regions through the constructed bridges in North Dakota. J. Bridge Eng. **15**(5), 542–552 (2009)

13. Huang, R., Mao, I., Lee, H.: Exploring the deterioration factors of RC bridge decks: a rough set approach. Comput.-Aided Civ. Infrastruct. Eng. **25**(7), 517–529 (2010)
14. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
15. Chawla, N.V.: Data mining for imbalanced datasets: an overview. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 875–886. Springer, Boston (2009). https://doi.org/10.1007/0-387-25465-X_40
16. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep Learning. The MIT Press, Cambridge (2016)