SIMULATION-BASED DATA-DRIVEN DAMAGE DETECTION FOR HIGHWAY BRIDGE SYSTEMS

X. Liang¹, K. Mosalam² and S. Muin³

ABSTRACT

Highway bridges are one of the most critical components in transportation infrastructure systems. Accumulated internal and concealed damages, due to aging or extreme events (e.g. earthquakes), make highway bridges vulnerable and pose a threat to the resiliency of local community. Therefore, these damages should be detected through structural health monitoring (SHM) algorithms at an early stage. Based on nonlinear time history analysis simulations on the investigated bridge system, a data-driven damage detection approach is explored on bridge columns, the most critical components of bridge systems. The paper starts with presenting damage feature selection where their effectiveness is demonstrated through unsupervised learning. The support vector machine, one of the representative supervised learning algorithms, is applied to several classification problems of engineering interests. Very promising results (estimation accuracies) are observed on both binary (damage vs. no damage, collapse vs. non-collapse) and multi-class (damage severity) classifications utilizing support vector machine.

¹Postdoctoral Scholar, Dept. of CEE, UC Berkeley, CA 94720 (email: benliangxiao@berkeley.edu)
²Taisei Professor of Civil Engineering and Director of PEER, Dept. of CEE, UC Berkeley, CA 94720
³Graduate Student Researcher, Dept. of CEE, UC Berkeley, CA 94720

Simulation-Based Data-Driven Damage Detection for Highway Bridge Systems

X. Liang¹, K. Mosalam² and S. Muin³

ABSTRACT

Highway bridges are one of the most critical components in transportation infrastructure systems. Accumulated internal and concealed damages, due to aging or extreme events (e.g., earthquakes), make highway bridges vulnerable and pose a threat to the resiliency of local communities. Therefore, these damages should be detected through structural health monitoring (SHM) algorithms at an early stage. Based on nonlinear time history analysis simulations on the investigated bridge system, a data-driven damage detection approach is explored on bridge columns, the most critical components of bridge systems. The paper starts with presenting damage feature selection where their effectiveness is demonstrated through unsupervised learning. The support vector machine, one of the representative supervised learning algorithms, is applied to several classification problems of engineering interests. Very promising results (estimation accuracies) are observed on both binary (damage vs. no damage, collapse vs. non-collapse) and multi-class (damage severity) classifications utilizing support vector machine.

Introduction

Reinforced concrete (RC) highway bridge systems are one of the most critical components in transportation infrastructure systems for transporting goods and people around natural terrains. Therefore, they are expected to sustain minor damage and maintain their functionality in the aftermath of earthquakes where accumulated damages should be detected through structural health monitoring (SHM) approaches. Currently, there are no quantifiable methods to determine if bridges are safe for reoperation after an extreme event [1]. However, advances in remote sensing, computing technologies and data analytics (e.g., machine learning techniques) offer the possibility of automating the structural health monitoring (SHM) process to assess and quantify the condition of structures in near-real time while removing the need for the intervention of human experts as far as possible. Instead of treating SHM as an inverse problem, this paper presents a data-driven

¹Postdoctoral Scholar, Dept. of CEE, UC Berkeley, CA 94720 (email: benliangxiao@berkeley.edu)
²Taisei Professor of Civil Engineering and Director of PEER, Dept. of CEE, UC Berkeley, CA 94720
³Graduate Student Researcher, Dept. of CEE, UC Berkeley, CA 94720

damage detection method based on acceleration data from more than 16,000 nonlinear time history analysis (NTHA) simulations on a RC highway bridge system under bidirectional ground motion input. The proposed method utilizes cumulative absolute velocity (CAV) [2] and Arias Intensity (\(I_A\)) of simulated acceleration data as damage features. Independent component analysis in addition to kernel density estimation are utilized to estimate joint probability density function (PDF) of damage features for normal conditions of investigated bridge system. The comparisons between the joint PDF estimated by this unsupervised learning algorithm and three groups of NTHA simulations, respectively corresponding to the earthquake scenarios with 50%, 10% and 2% probability of exceedance (POE) in 50 years, show that relative CAV and relative \(I_A\) are effective damage indicators for bridge columns, the most critical components of RC highway bridge systems. Support vector machine (SVM), one of the representative supervised learning algorithms, is applied subsequently on these damage features to identify occurrence of collapse and existence as well as extent of the damage.

**Computational Bridge Model and Ground Motion Selection**

A representative RC highway bridge system, *Jack Tone Road Overcrossing*, is selected for conducting this study. The selected bridge, designed after 2000 with characteristics and configuration summarized in Fig. 1, reflects common bridge engineering practice in California. Extensive analytical modeling and simulations for this bridge can be found in [3]. The software OpenSees [4] is used for the simulations. The models explicitly include seat-type abutments (seat length = 33.85 in.), shear keys, expansion joints, column-bents, and superstructure. The used concrete has compressive strength \(f'_c = 5.0\) ksi and modulus of elasticity \(E_c = 4030.5\) ksi. The used reinforcing steel follows ASTM A706 standards. The adopted modeling is briefly described in the following paragraphs.

Caltrans SDC [5] requires the superstructure of a bridge to be capacity protected and accordingly to remain elastic. Thus, the bridge superstructure that consists of the bridge deck and the cap-beam is modeled with elastic beam-column elements using uncracked section properties. The cap-beam is assigned arbitrarily high torsional and out-of-plane stiffness since the cap-beam and the deck are integrally constructed. The mass of the superstructure, including the rotational mass, is distributed to the superstructure elements to accurately capture the dynamic response. The bridge column is modeled with nonlinear force-based beam-column elements using fiber-discretized cross-sections and 10 integration points along the column height. Three constitutive models are utilized simultaneously in a fiber discretized cross-section: (1) confined concrete for the core, (2) unconfined concrete for the cover, and (3) steel for the reinforcing bars. In OpenSees, *Concrete01* constitutive model is a uniaxial Kent-Scott-Park concrete material object with degraded linear unloading/reloading stiffness and no tensile strength. It is used for both the cover and core concrete according to the model developed by Mander et al. [6]. The steel reinforcing bars are modeled by *Steel02* material, a uniaxial Giuffre-Menegotto-Pinto steel material object with isotropic strain hardening, with the steel modulus of elasticity and the expected yield strength are set as 29,000.0 ksi and 68.0 ksi, respectively.
The abutment modeling approach explicitly takes into account the longitudinal response of the backfill and expansion joint, the transverse response of the shear keys, and the vertical response of the bearing pads and the stemwall. Two nonlinear springs, one at each end, connected in series to gap elements, are used to model the passive backfill response and the expansion joint [7], respectively. The shear key response is modeled using an elastic-perfectly-plastic backbone relationship. The vertical response of the bearing pads and stemwall is modeled by two parallel springs, one at each end (note that only one side is labelled in Fig. 1), to represent the stiffness values.

Based on a magnitude 7 earthquake scenario in [8], 99 pairs of seed bidirectional horizontal GM records are selected from the PEER Next Generation Attenuation (NGA) Project GM database [9]. These 99 pairs of GM records are subsequently scaled based on the distribution of PGV from [10]. In this study, 25 values of PGV are selected to represent this distribution. Moreover, different intercept angles, varying from 0° to 150° with an increment of 30°, are investigated (Fig. 1). Therefore, considering the above-mentioned six intercept angles for all 99 unscaled GMs with 25 values of PGV, \(6 \times 99 \times 25 = 14,850\) NTHA simulations are performed in total where simulations of the first five intercept angles are used for training and those of the last intercept angle are used as a testing set.

**Damage Feature Extraction**

An ideal damage feature is certain low-dimensional quantity extracted from the system response data that is highly sensitive to the damage condition of the structure. It is to be noted that these features are the quantities that the pattern recognition and machine learning algorithms will subsequently analyze in an effort to identify and quantify the damage. Therefore, these damage features are preferably to change in a monotonic fashion with increasing damage levels. The damage features investigated in this paper are cumulative absolute velocity (CAV) and Arias
Intensity \((I_A)\) of simulated acceleration data from top and bottom (i.e. input GM) of the bridge column. The \(CAV\) and \(I_A\) are defined as follows (Fig. 2):

\[
CAV = \int_0^{T_d} |a(t)| \, dt
\]

\[
I_A = \left(\frac{\pi}{2g}\right) \int_0^{T_d} a^2(t) \, dt
\]

where \(a(t)\) is the acceleration time history and \(T_d\) is the duration of the earthquake. Therefore, both features account for the contributions from the amplitude and duration. To access the damage conditions of bridge column, relative \(CAV\) and relative \(I_A\) are introduced as follows:

\[
R_{CAV} = \frac{CAV^{ct}}{CAV^{input}}
\]

\[
R_{I_A} = \frac{I_A^{ct}}{I_A^{input}}
\]

where \(CAV^{ct}\) and \(I_A^{ct}\) are respectively \(CAV\) and \(I_A\) of the bridge column top acceleration time history, and \(CAV^{input}\) and \(I_A^{input}\) represent \(CAV\) and \(I_A\) of the input GM. These two ratios tend to be decreasing with increasing energy dissipation and thus indicating increasing levels of acquired damages on the bridge column. Considering the bidirectional GM input, totally four damage features are taken into account.

![Cumulative absolute velocity and Arias intensity](image)

**Figure 2.** Cumulative absolute velocity and Arias intensity.

**Pattern Recognition**

A pattern recognition (PR) algorithm is one that assigns a class label to a sample of measured data, usually from a finite set. Many PR algorithms work by training a diagnostic. In the context of data analytics (especially machine learning), there are two general types of learning: unsupervised and supervised learning. Both types are investigated in this study.
Unsupervised Learning

The idea of unsupervised learning is that only trained data from the undamaged condition of the structure or system will be used to establish the diagnostic. A statistical model (e.g. joint probability density function) of damage features for the undamaged condition is subsequently created and data acquired during monitoring are compared with this model. In this paper, independent component (IC) analysis in addition to the kernel density estimation (KDE) [11] are utilized to develop this joint PDF based on the selected 99 GMs with small scaling factors ranging from 0.01 to 0.1 with an increment of 0.01, i.e. totally 990 NTHA simulations representing the undamaged conditions of investigated bridge system. The procedures are as follows:

1. Apply IC analysis to the data $\mathbf{X}$ in the original space to obtain the data $\mathbf{Y}$ in the IC space as follows:

$$\mathbf{Y} = \mathbf{P}(\mathbf{X} - \mu_X)$$
$$\mathbf{X} = \mu_X + \mathbf{QY} \quad (5)$$

where $\mu_X$ stands for mean matrix of $\mathbf{X}$ and $\mathbf{Q} = \mathbf{P}^{-1}$.

2. Apply KDE to each IC and the joint PDF in the IC space can be obtained as follows:

$$f_Y(y) = f_{y_1}(y_1) \cdot f_{y_2}(y_2) \cdots f_{y_n}(y_n) \quad (6)$$

3. The joint PDF in the original space can be obtained as follows:

$$f_X(x) = \frac{1}{|\mathbf{Q}|} f_Y[\mathbf{P}(\mathbf{x} - \mu_X)] \quad (7)$$

Figure 3. The comparisons between joint PDF of $R_{CAV}$ for the undamaged condition and $R_{CAV}$ of three damage levels: 50% (left), 10% (middle), and 2% (right) POE in 50 years.

Moreover, three groups of 40 GMs [12], which are corresponding to the earthquake scenarios with 50%, 10% and 2% probability of exceedance (POE) in 50 years, are selected to represent three damage levels of the bridge. Figs. 3 and 4 respectively show the comparisons between the joint PDF and the three groups for $R_{CAV}$ and $R_{IA}$, respectively. Clear monotonic trends are observed for both ones with increasing damage. Therefore, they are effective damage
indicators for bridge column of the investigated RC highway bridge system and are ready to be used as damage features in supervised learning algorithm introduced in the next subsection.

Figure 4. The comparisons between joint PDF of $R_{t_a}$ for the undamaged condition and $R_{t_a}$ of three damage levels: 50% (left), 10% (middle), and 2% (right) POE in 50 years.

Supervised Learning

This type of learning algorithm, in which the diagnostic is trained by showing it the true label for each data set, is a necessity to access the type and severity of the damage. Data from different damage states will be used to train or classify. In this study, support vector machine (SVM), one of the representative supervised learning algorithms, is investigated. The idea of the SVM is to construct a hyperplane in Eq. 8 to separate two different classes of data samples ($y_i \in \{-1, 1\}$) and to maximize the margin from the hyperplane to the closest data points in either class.

$$f(x) = h(x)^T \beta + \beta_0 = 0$$

(8)

This hyperplane is in terms of the extended features

$$h(x) = (h_1(x), h_2(x), \ldots, h_M(x))^T$$

(9)

where $M$ is the total number of selected basis functions. Accordingly, the optimization problem for SVM can be expressed as follows [13]:

$$\min_{\beta, \beta_0} \frac{1}{2} ||\beta||^2 + C \sum_{i=1}^{N} \xi_i$$

subject to $\xi_i \geq 0, \ y_i(h(x_i)^T \beta + \beta_0) \geq 1 - \xi_i, \ \forall i$

(10)

where $N$ is the total number of sampled points, $C$ is the cost parameter to control the tradeoff of bias and variance, and $\xi$ is the slack variable to allow for some data points to be on the wrong side of margin. Eq. 8 can be further written as
\[ f(x) = h(x)\beta + \beta_0 \]
\[ = \sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + \beta_0 \]  
(11)

where \( K(x, x') \) is the kernel function, which is the inner product of \( h(x) \) and \( h(x') \), i.e.

\[ K(x, x') = \langle h(x), h(x') \rangle \]  
(12)

In this paper, the radial basis kernel function in Eq. 13 is used.

\[ K(x, x') = \exp\left(-\gamma\|x - x'\|^2\right) \]  
(13)

The two most important hyperparameters \( C \) and kernel scale \( \gamma \) are determined through Bayesian optimization [14] to minimize the cross-validated loss of the SVM model. For illustration, Fig. 5 shows the training results using \( R_{CAV} \) and \( R_{IA} \) respectively as damage features for occurrence of collapse (i.e. peak column drift ratio exceeds 8% or not [15]) and existence of damage (i.e. peak column drift ratio exceeds 2% or not) where the training and testing accuracies are documented in Table 1. The same two classification problems are also conducted considering four damage features, i.e. \( R_{CAV} \) and \( R_{IA} \) in both \( x \) and \( y \) directions. Improved accuracies observed on both training and testing sets, as shown in Table 1, comparing to those in Fig. 5 (e.g. 90.30% vs. 83.49%). The SVM can also be extended to conduct multi-class classification problems. This paper provides a three-class classification: no damage, damage without collapse, and collapse, i.e. peak column drift ratio is below 2%, between 2% and 8%, and above 8%. Very promising accuracies of \( \sim 80\% \) are achieved (Table 1). Also, Figs. 6-8 give the confusion matrices of training and testing sets for the three cases where the predicting accuracies of each class are provided. It is noted that the accuracies for training and testing sets for all cases are comparable, indicating that the SVM classifiers achieve a good robustness on the tradeoff of bias and variance.

![Figure 5. SVM training results for occurrence of collapse using \( R_{CAV} \) (left) and existence of damage using \( R_{IA} \) (right).](image-url)
Figure 6. Confusion matrix for binary classification: damaged vs. no damage.

Figure 7. Confusion matrix for binary classification: collapsed vs. non-collapse.

Figure 8. Confusion matrix for multi-class classification: severity of damage.
Table 1. Training and testing accuracies for investigated classification cases.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Damage Features</th>
<th>Training Accuracy</th>
<th>Testing Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrence of collapse</td>
<td>$R_{CAV}$ (x, y)</td>
<td>83.49%</td>
<td>83.22%</td>
</tr>
<tr>
<td>Existence of damage</td>
<td>$R_{IA}$ (x, y)</td>
<td>85.62%</td>
<td>86.62%</td>
</tr>
<tr>
<td>Occurrence of collapse</td>
<td>$R_{CAV}$ &amp; $R_{IA}$ (x, y)</td>
<td>90.30%</td>
<td>88.11%</td>
</tr>
<tr>
<td>Existence of damage</td>
<td>$R_{CAV}$ &amp; $R_{IA}$ (x, y)</td>
<td>87.57%</td>
<td>87.78%</td>
</tr>
<tr>
<td>No damage, damaged without collapse, and collapsed</td>
<td>$R_{CAV}$ &amp; $R_{IA}$ (x, y)</td>
<td>81.17%</td>
<td>77.75%</td>
</tr>
</tbody>
</table>

Conclusions

This paper presents a data-driven damage detection approach for bridge columns based on NTHA simulations on the investigated RC highway bridge system. This SHM process mimics putting four accelerometers on the bridge column, of which two are measuring the bidirectional input GM excitation and the other two are recording the acceleration time histories of the column top in both longitudinal and transverse directions. The $R_{CAV}$ and $R_{IA}$ are extracted from the acceleration time histories and are proved to be good damage features through unsupervised learning. The SVM, as one of the representative supervised learning algorithms, is applied to several binary and multi-class classification problems of engineering interests: occurrence of collapse, existence and severity of damage. Very promising accuracies and robustness are observed which offers the possibility to utilize on board computing of sensors for near-real time damage assessment. Future work includes applications on more RC highway bridge systems, putting more sensors (e.g. two more accelerometers at the middle height of the column), and further improvement of the training and testing accuracies (e.g. putting more weights on misclassified data points).

Acknowledgments

Professor K.M. Mosalam is a core-principal investigator of Tsinghua-Berkeley Shenzhen Institute (TBSI). The authors acknowledge the funding support from TBSI. Prof. Farzin Zareian, University of California, Irvine, is acknowledged for providing the OpenSees model of the analyzed bridge.

References

4. McKenna F, Fenves GL, Filippou FC. The Open System for Earthquake Engineering Simulation, University of


10. Campbell KW, Bozorgnia Y. NGA Ground Motion Model for the Geometric Mean Horizontal Component of PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10s. *Earthquake Spectra* 2008; **24**(1): 139-171.


