A data-driven approach for sensor data reconstruction for bridge monitoring

* Kincho H. Law [1], Seongwoon Jeong[2], Max Ferguson[3]

[1),2),3)] *Dept. of Civil and Environ. Eng., Stanford University, Stanford, CA 94305, USA*
[1)] *law@stanford.edu*

ABSTRACT

The purpose of this paper is to explore the potential use of machine learning techniques to build data-driven models for diagnostic analysis of civil infrastructures. The discussion will focus on the reconstruction of sensor data collected from a bridge monitoring system using Support Vector Regression (SVR). The methodology can be used for constructing missing data from faulty sensors, as well as for checking the integrity of the sensor data, thereby potentially revealing issues, if any, on a target structure. The proposed approach is validated using the sensor data sets acquired from the Telegraph Road Bridge located in the state of Michigan.

1. INTRODUCTION

The purpose of bridge monitoring is to ensure that a target bridge structure is in a healthy and reliable state. Sensor data provides valuable information about the structural state of a bridge as well as to detect the onset of damage and to understand the long-term health of a bridge. The sensor data collected, however, may be corrupted or erroneous due to various causes, such as faulty sensor, battery problem and incomplete data transmission. Problematic sensors and erroneous or missing data could lead to incorrect bridge management decision. A procedure that can validate the correctness of the sensor data and that possibly construct missing data can be valuable to ensure the integrity of the monitoring system.

There have been research efforts targeted to reconstruct missing data collected from infrastructure monitoring systems. A Kalman filter-based approach has been proposed to predict missing data based on the state vector (Cipar and Romera 1997). For example, Kim *et al*. (2011) adopt the Kalman filter algorithm to recover missing vibration data measured from a bridge monitoring system. Compressive sensing has

---

[1)] Professor
[2)] Graduate student
[3)] Graduate student

also been used to cope with missing data due to the packet loss during data transmission (Zou *et al*. 2014). Most approaches have been focused on the information of a single sensor. In this study, we explore the possibility of validating sensor data or reconstructing missing data utilizing multiple sensors in the bridge monitoring system.

In this paper, we explre the potential use of machine learning techniques to build data-driven models for diagnostic analysis of civil infrastructures with an emphasis on the reconstruction of sensor data. We adopt Support Vector Regression (SVR) algorithm for the data reconstruction. The reconstructed data can be used not only for constructing missing data from faulty sensors, as well as for cross checking the integrity of the sensor data. The proposed method is tested using the sensor data sets acquired from the Telegraph Road Bridge located in Monroe, Michigan.

2. EPSILON SUPPORT VECTOR REGRESSION

The proposed sensor data reconstruction approach aims to build a data-driven model that predicts the data of a sensor based on the data collected by other sensors. In this exploratory study, we adopt Support Vector Regression (SVR). Detailed description of SVR can be found in Schölkopf and Smola (2002). Here, we briefly summarize the formulation of Epsilon Support Vector Regression (ϵ-SVR) for the sensor data reconstruction.

The target sensor data is described as a simple linear regression function $\mathrm{f}(x, w)$ of sensor data sets x and regression weights w as:

$$\mathrm{f}(x, \mathrm{w}) = w^T x + b \tag{1}$$

where b is the intercept term of the linear regression. One way to find the parameters w and the coefficients b is to minimize the error E over the training data set P. Here, the error E is defined as:

$$\mathrm{E} = \sum_{\mathrm{p} \in \mathrm{P}} \frac{1}{2} \left(\mathrm{f}(x_p, w) - y_p \right)^2 \tag{2}$$

where y_p is the measured value of the target sensor data. An alternative approach that can provide more modeling flexibility is the Epsilon Support Vector Regression (ϵ-SVR) in which an allowable error ϵ is specified as (Vapnik, 1995):

$$|y - \mathrm{f}(x, w)|_\epsilon = \max(0, |y - \mathrm{f}(x, w)| - \epsilon\ ) \tag{3}$$

Eq (3) implies that only the values beyond the specified margin ϵ will be penalized. By introducing ϵ, the loss function for training set is defined as:

$$L(w, b) = \frac{1}{2} \left| |w| \right|^2 + C \sum_{i=1}^{n} |y_i - f(x_i, w)|_\epsilon \tag{4}$$

where C is termed regularization parameter. For simplification, slack variables ξ and ξ^* are introduced as follows:

$$\xi_i^* = |y_i - f(x_i, w)| - \varepsilon \text{ for data 'above' an } \varepsilon\text{-tube} \quad (5)$$

$$\xi_i = |y_i - f(x_i, w)| - \varepsilon \text{ for data 'below' an } \varepsilon\text{-tube} \quad (6)$$

The minimization of the loss function shown in Eq (4) can be state as:

$$\min_{\boldsymbol{w},b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (7)$$

where

$$\begin{gathered} y_i - w^T x_i - b \le \epsilon + \xi_i, \quad i = 1, \dots, n \\ w^T x_i + b - y_i \le \epsilon + \xi_i^*, \quad i = 1, \dots, n \\ \xi_i,\ \xi_i^* \ge 0 \quad i = 1, \dots, n \end{gathered}$$

The optimization problem can be expressed as a Lagrangian dual problem. The objective is to find the model parameters w and b, which can then be used to reconstruct sensor data.

3. APPLICATION TO SENSOR DATA RECONSTRUCTION

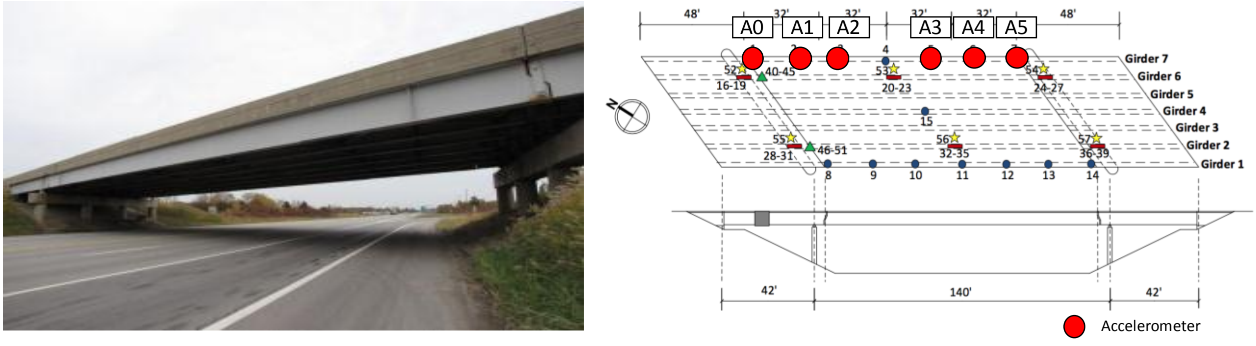
To illustrate the sensor data reconstruction, we use the acceleration data sets collected from the wireless sensor network on the Telegraph Road Bridge (Fig. 1(a)) located in the state of Michigan (Zhang *et al*. 2016). Accelerometers are installed along *Girder 1* and *Girder 7* of the bridge as shown in Fig 1(b). The accelerometers measure the bridge vibration for one minute duration every 2 hours with 200Hz sampling rate. In this study, we use the acceleration data collected from the six accelerometers labeled as *A0* to *A5* along *Girder 1* of the bridge.

We select 80,000 sequentially sampled acceleration data from each of the sensors. The data are collected over 400 second period. The first half (i.e., 40,000 data points measured in the first 200 seconds) of the data set are assigned as the training data set, while the second half of the data set are assigned as the testing data set as summarized below.

$$\text{Training data set: } \boldsymbol{x}^{(i)} = \left[x_0^{(i)}, x_1^{(i)}, \dots, x_5^{(i)}\right] \quad i = 1, \dots, 40{,}000 \quad (8)$$

$$\text{Testing data set: } \boldsymbol{x}^{(i)} = \left[x_0^{(i)}, x_1^{(i)}, \dots, x_5^{(i)}\right] \quad i = 40{,}001, \dots, 80{,}000 \quad (9)$$

where $x_j^{(i)}$ denotes the acceleration measurement data by sensor j at time step i. Figure 2 shows the sensor data sets including both the training set and the testing set.



(a) Side view (b) Sensor layout

Fig. 1 Telegraph road bridge (Monroe, Michigan)

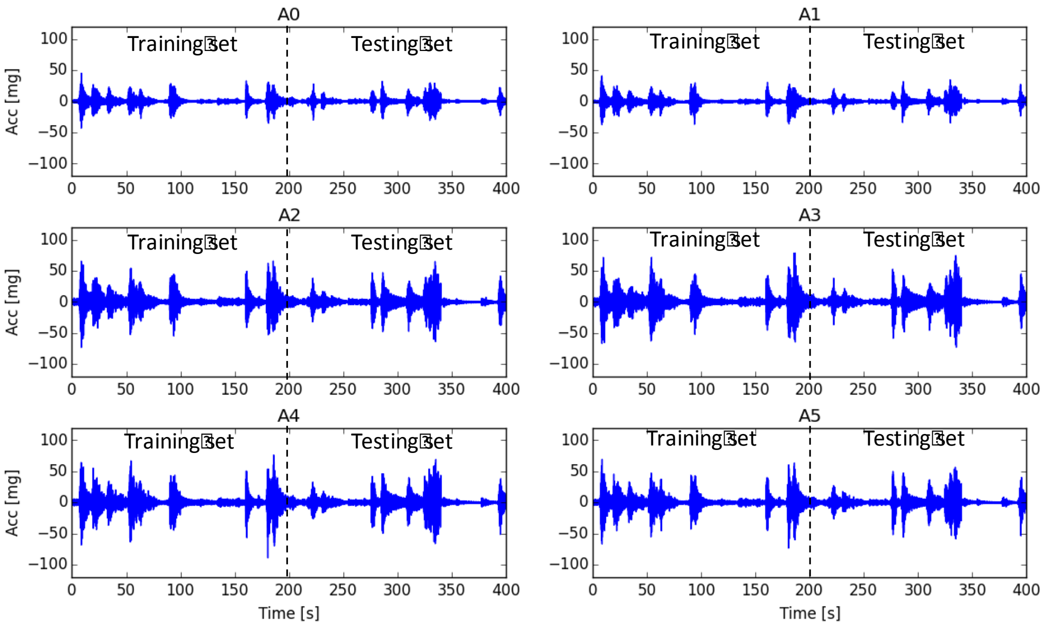


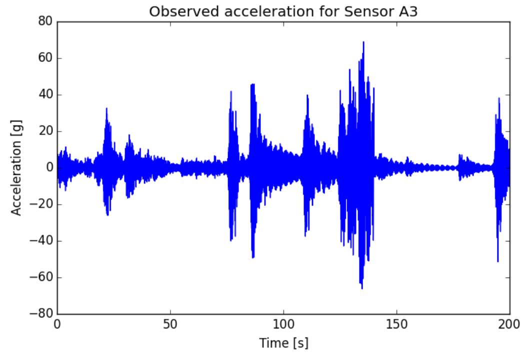
Fig. 2 Training set and testing set of acceleration data

We select the accelerometer A3 located at the mid-span as the target sensor whose data is to be reconstructed by ϵ-SVR. To construct the acceleration data for accelerometer A3, the input $\boldsymbol{x}^{(i)}$ and the target feature $y^{(i)}$ are as follows.

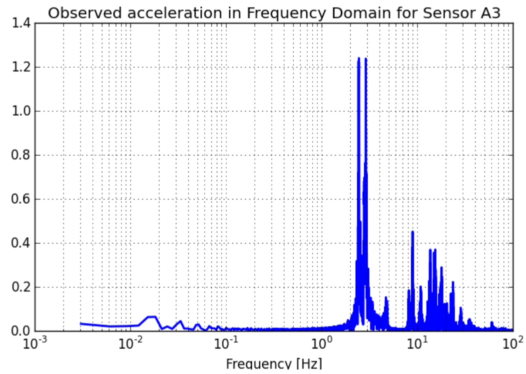
$$\boldsymbol{x}^{(i)} = \left[x_0^{(i)}, x_1^{(i)}, x_2^{(i)}, x_4^{(i)}, x_5^{(i)}\right] \tag{10}$$

$$y^{(i)} = x_3^{(i)} \tag{11}$$

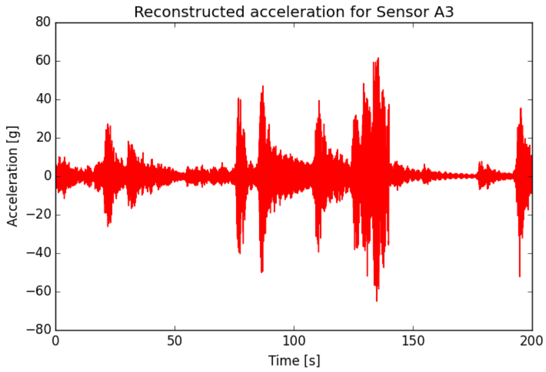
Given the input data and the target feature, we first train the ϵ-SVR model for prediction target $y^{(i)}$ based on the input $\boldsymbol{x}^{(i)}$ by solving the optimization problem described in Eq (7) using the training data set. Once the model is built, we construct $y^{(i)}$ using the input data $\boldsymbol{x}^{(i)}$ from testing data set based on the ϵ-SVR model. Here, we use scikit-learn (http://scikit-learn.org/), a machine learning package for Python, to perform training and testing (Pedregosa *et al*. 2011). Fig. 3 compares the measured acceleration data and the reconstructed acceleration in both the time and frequency domains. By comparing the measured acceleration data shown in Figure 3(a) and the reconstructed acceleration data shown in Figure 3 (c), it can be seen that the ϵ-SVR model is able to reconstruct the sensor data with very good precision. Furthermore, as can be seen from the Fourier spectra of the measured and reconstructed data, shown in Figure 3(b) and 3(d), respectively, dominant frequencies obtained from reconstructed data match well with the measurement data.



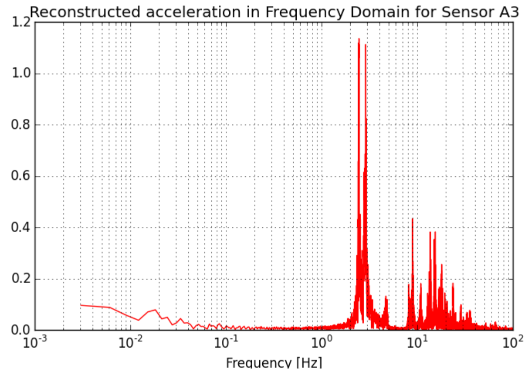
(a) Observed acceleration



(b) Observed acceleration (Fourier spectra)



(c) Reconstructed acceleration



(d) Reconstructed acceleration (Fourier spectra)

Fig. 3 Comparison of observed acceleration and the reconstructed acceleration

4. DISCUSSION

In this paper, we explore the potential use of machine learning techniques to build data-driven models for bridge monitoring applications. The discussion focuses on the reconstruction of sensor data using the ϵ-SVR algorithm. The application results show that the proposed approach can reconstruct the sensor data with very good precision. Continuing research will be conducted to extend this preliminary finding for cross validating the integrity of the sensor data, thereby potentially revealing issues on a target structure.

ACKNOWLEDGMENTS

This research is supported by a Grant No. 13SCIPA01 from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea government and Korea Agency for Infrastructure Technology Advancement (KAIA). The research is also partially supported by a collaborative project funded by the US National Science Foundation (Grant No. ECCS-1446330 to Stanford University and Grant No. ECCS-1446521 to the University of Michigan). The authors thank the Michigan Department of Transportation (MDOT) for access to the Telegraph Road Bridge and the Newburg Road Bridge and for offering support during installation of the wireless monitoring system. Any opinions, findings, conclusions or recommendations expressed in this paper are solely those of the authors and do not necessarily reflect the views of NSF, MOLIT, MDOT, KAIA or any other organizations and collaborators.

REFERENCES

Cipra, T. and Romera, R. (1997), “Kalman filter with outliers and missing observations,” *Test*, **6**(2), pp.379-395.

Kim, C.W., Kawatani, M., Ozaki, R. and Makihata, N. (2011), “Recovering missing data transmitted from a wireless sensor node for vibration-based bridge health monitoring,” *Structural Engineering and Mechanics*, **38**(4), pp.417-428.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J. (2011), “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, **12**(Oct), pp.2825-2830.

Schölkopf, B. and Smola, A.J. (2002), “Learning with kernels: support vector machines, regularization, optimization, and beyond,” MIT press.

Vapnik, V. (1995), “The nature of statistical learning theory,” Springer.

Zhang, Y., O’Connor, S.M., van der Linden, G., Prakash, A. and Lynch, J.P. (2016), “SenStore: A Scalable Cyberinfrastructure Platform for Implementation of Data-to-Decision Frameworks for Infrastructure Health Management,” *Journal of Computing in Civil Engineering*, **30**(5), 04016012.

Zou, Z., Bao, Y., Li, H., Spencer, B.F. and Ou, J. (2015), “Embedding compressive sensing-based data loss recovery algorithm into wireless smart sensors for structural health monitoring,” *IEEE Sensors Journal*, **15**(2), pp.797-808.