LETTER

Long Short-Team Memory for Forecasting Degradation Recovery Process with Binary Maintenance Intervention Records

Katsuya KOSUKEGAWA[†], ***Nonmember and*** **Kazuhiko KAWAMOTO**[†a)], ***Member***

SUMMARY We considered the problem of forecasting the degradation recovery process of civil structures for prognosis and health management. In this process, structural health degrades over time but recovers when a maintenance intervention is performed. Maintenance interventions are typically recorded in terms of date and type. Such records can be represented as binary time series. Using binary maintenance intervention records, we forecast the process by using Long Short-Term Memory (LSTM). In this study, we experimentally examined how to feed binary time series data into LSTM. To this end, we compared the concatenation and reinitialization methods. The former is used to concatenate maintenance intervention records and health data and feed them into LSTM. The latter is used to reinitialize the LSTM internal memory when maintenance intervention is performed. The experimental results with the synthetic data revealed that the concatenation method outperformed the reinitialization method.

key words: *degradation recovery process, maintenance intervention, time series forecasting, long short-term memory*

1. Introduction

Civil structures, such as railroad tracks, degrade over time owing to the wear, corrosion, rupture, and deformation of components. To maintain such structures, human operators monitor their health and repair them if necessary. Forecasting the structure health is critical for project and health management [1], [2] rather than reactive management because the operators can determine whether maintenance interventions are necessary, in advance.

For prognostics and health management, we consider the problem of forecasting a degradation recovery process, as shown in Fig. 1 (top). In this process, the structure health degrades over time but recovers when maintenance intervention is performed. Maintenance interventions are typically recorded in terms of date and type. In this study, we represented such records as a binary time series such as $[0, 0, 1, 0, 0, \ldots]$, indicating that maintenance intervention was performed on the third day.

We propose a long short-term memory model (LSTM) [3] to predict the degradation recovery process using binary maintenance intervention records. There are several possible ways to feed such binary time series data into LSTM. However, these input methods have not been sufficiently explored. We experimentally compared two methods for inputting the maintenance intervention records in LSTM using synthetic data. One is the concatenation of maintenance intervention records and health data and feeding them into LSTM [4]. A concatenation method is a straightforward approach to deep learning. The other is to re-initialize the LSTM internal memory when performing maintenance intervention. This method re-initializes LSTM each time some maintenance intervention is performed. Thus, LSTM can focus on forecasting degradation processes. Experimental results using synthetic data exhibit that the first concatenation method outperformed the second initialization method. Furthermore, an ablation study with various observation noise levels revealed that both methods provide better performance than LSTM without the maintenance intervention records. The main contributions of this study are summarized as follows:

- We revealed that the concatenation method outperforms the initialization method in forecasting performance.
- We demonstrated that the binary maintenance intervention records contribute to improving forecasting performance. This result shows that LSTM can be learned from binary time series data.

2. Method

In this section, we formulate the problem of forecasting the degradation recovery process with maintenance interventions. Then, we describe the proposed methods for inputting maintenance intervention data into LSTM.

2.1 LSTM Model for Forecasting

We denote the health of the structure at time t as $x_t^* \in [0, 1]$, where $x_t^* = 1$ indicates the best health and a lower x_t^* indicates worse health. We assume that the operators observe noisy health x_t as $x_t = x_t^* + \varepsilon_t$, where x_t^* is the ground truth and ε_t is Gaussian noise with zero mean.

This study considered two maintenance interventions: partial repair and replacement [5]. If either a partial repair or replacement is performed, the health of the structure is restored. The difference between the two interventions was that the degradation rate after the replacement was lower than that after partial repair because replacement refers to the replacement of an entire structure with a new one. We denote partial repair and replacement at time t as $p_t, r_t \in \{0, 1\}$, respectively, where p_t and r_t take one exclusively, and $p_t = 1$ or $r_t = 1$ indicates that the intervention was performed.

We aimed to build an LSTM based forecasting model

Manuscript received March 6, 2023.
Manuscript revised July 12, 2023.
Manuscript publicized August 7, 2023.

[†]The authors are with Chiba University, Chiba-shi, 263-8522 Japan.
a) E-mail: kawa@faculty.chiba-u.jp

DOI: 10.1587/transfun.2023EAL2019

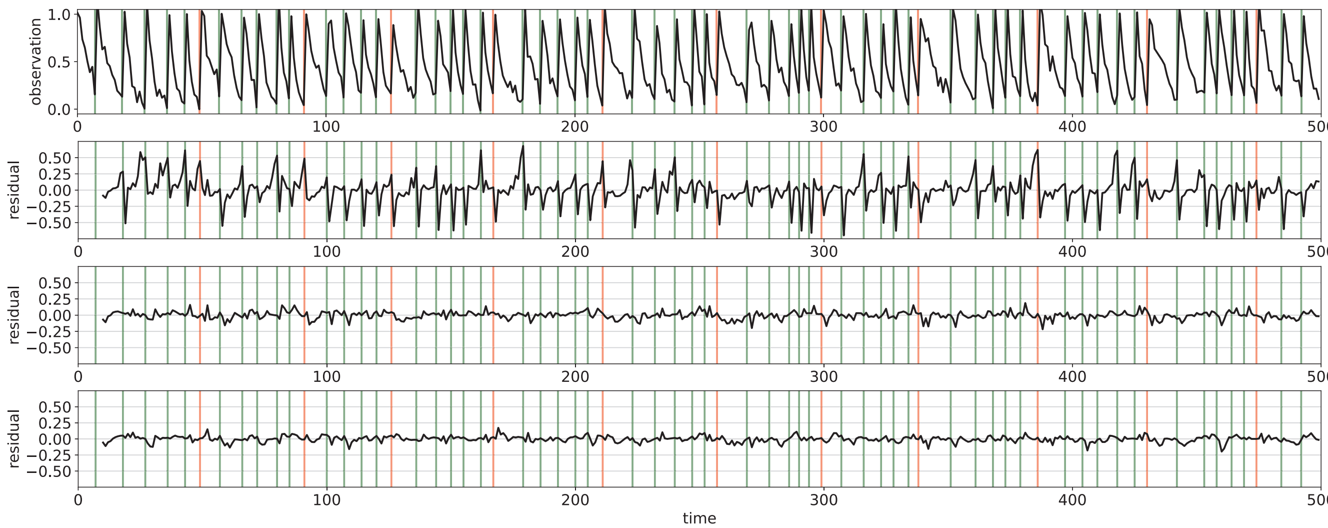

Fig. 1 Example of experimental results. The top figure shows a time series of observations for the degradation recovery process. The other figures show the residuals between the ground truth and the predictions by (second) the vanilla LSTM, (third) the reinitialization method, and (bottom) the concatenation method, respectively.

$\hat{x}_{t+1} = f(x_{t-T+1:t}, p_{t-T+1:t}, r_{t-T+1:t}; \theta)$, where we denote $x_{t-T+1:t} = \{x_{t-T+1}, \ldots, x_t\}$, $p_{t-T+1:t} = \{p_{t-T+1}, \ldots, p_t\}$, and $r_{t-T+1:t} = \{r_{t-T+1}, \ldots, r_t\}$, respectively, and θ is parameters of the LSTM. We set $T = 10$ during the experiments. To determine the optimal parameter $\hat{\theta}$, we minimize the squared loss function with respect to θ:

$$L(\theta) = \frac{1}{N - T + 1} \sum_{t=T}^{N} (x_{t+1} - \hat{x}_{t+1})^2 \rightarrow \min, \tag{1}$$

where N is the length of time series data.

2.2 Inputting Binary Maintenance Intervention Records

We examined two methods for inputting maintenance intervention records p_t, r_t into the LSTM: concatenation and reinitialization.

Concatenation: This method concatenates the health observation x_t and maintenance records p_t, r_t as $\tilde{x}_t = (x_t, p_t, r_t)$. The concatenated vector $\tilde{x}_t$ is inputted into the first layer of LSTM. Although the LSTM is not explicitly trained to estimate the recovery timing and the degradation rate, we expect that the LSTM can capture degradation-recovery patterns as part of its internal representations owing to its ability to learn temporal dependencies.

Reinitialization: This method reinitializes the memory cell c_t and hidden state h_t of LSTM when a partial repair ($p_t = 1$) or replacement ($r_t = 1$) is performed as follows:

$$h_{t+1} = \begin{cases} h'_{t+1}(1 - p_t) + h^{(\mathrm{p})}_{\mathrm{init}} p_t & \text{for partial repair} \\ h'_{t+1}(1 - r_t) + h^{(\mathrm{r})}_{\mathrm{init}} r_t & \text{for replacement} \end{cases} \tag{2}$$

$$c_{t+1} = \begin{cases} c'_{t+1}(1 - p_t) + c^{(\mathrm{p})}_{\mathrm{init}} p_t & \text{for partial repair} \\ c'_{t+1}(1 - r_t) + c^{(\mathrm{r})}_{\mathrm{init}} r_t & \text{for replacement} \end{cases} \tag{3}$$

where h'_{t+1}, c'_{t+1} are the outputs of the vanilla LSTM cell, and $h^{(\mathrm{p})}_{\mathrm{init}}, h^{(\mathrm{r})}_{\mathrm{init}}, c^{(\mathrm{p})}_{\mathrm{init}}, c^{(\mathrm{r})}_{\mathrm{init}}$ are learnable reinitialization parameters. These parameters are learned during back-propagation for Eq. (1); the gradients $\partial L/\partial h^{(*)}_{\mathrm{init}}$ and $\partial L/\partial c^{(*)}_{\mathrm{init}}$, where $*$ indicates either p and r, are used in back-propagation when either intervention is performed. This reinitialization is intended to initialize the LSTM model each time the maintenance interventions were also administered. Thus, LSTM focuses on forecasting the degradation process.

3. Experiments

We experimentally compared the two methods for inputting maintenance intervention data into LSTM using the degradation recovery process.

3.1 Experiments Setting

We independently generate 18 time series data with $N = 500$ lengths using a degradation-recovery process based on geometric Brownian motion [6] described in Appendix. For random noise ε_t, we use normal distributions $N(0, \sigma^2)$ with $\sigma = 0.01, 0.03, 0.05, 0.07, 0.09$. We divide the 18 time series data into six for training, two for validation, and ten for testing. Figure 1 (top) shows a test time series data for $\sigma = 0.05$. In the figure, the green and red vertical lines indicate that partial repair and replacement were performed.

For training LSTMs, we used the Adam optimizer [7] with hyperparameters $(\beta_1, \beta_2) = (0.9, 0.999)$. We set the batch size to 16 and the number of epochs to 100. We use the mean squared error loss. Using Optuna [8], we optimized the following hyperparameters of LSTM: number of hidden units, number of layers, dropout rate, and learning rate.

Table 1 Comparison results of RMSE with the vanilla LSTM, the reinitialization method, and the concatenation method. The best RMSEs for each noise level are highlighted in bold.

noise level (σ)	0.01	0.03	0.05	0.07	0.09	0.11	0.13	0.15
vanilla LSTM	0.247	0.245	0.235	0.225	0.217	0.213	0.207	0.205
reinitialization	0.128	0.132	0.136	0.135	0.140	0.144	0.143	0.149
concatenation	**0.114**	**0.117**	**0.125**	**0.120**	**0.125**	**0.128**	**0.134**	**0.144**

3.2 Evaluation Metrics

As the evaluation metric, we used the root mean squared error (RMSE) as follows:

$$\mathcal{L} = \sqrt{\frac{1}{(N-T)M} \sum_{t=T+1}^{N} \sum_{m=1}^{M} \left(x_t^{*(m)} - \hat{x}_t^{(m)} \right)^2}, \tag{4}$$

where M is the number of time series data for testing, set to $M = 10$, and $x_t^{*(m)}$ and $\hat{x}_t^{(m)}$ are the ground truth and predictionat time t of mth time series data, respectively.

3.3 Experiment Results

Table 1 lists the RMSEs for vanilla LSTM, LSTM with reinitialization, and LSTM with concatenation. The vanilla LSTM was trained using only the health time series, while the other two methods use maintenance intervention records of partial repairs and replacements. Their results indicated that the concatenation method provided the best forecasting performance. Figure 1 show example results for the noise level $\sigma = 0.05$. The second, third, and bottom rows show the residuals between the ground truth and predictions by vanilla LSTM reinitialization and concatenation methods, respectively.

From Fig. 1 (second), vanilla LSTM provides poor predictions, especially at and after the interventions, indicated for green and red vertical lines, respectively; This was because the vanilla LSTM was not provided with maintenance intervention records. Contrary, from Fig. 1 (third and bottom), the other two methods provided better predictions than the vanilla LSTM; This indicates that binary records are useful for forecasting the jumps in the time series data. When comparing the reinitialization method with the concatenation method, the latter had smaller RMSEs for all noise levels, as shown in Table 1. From Fig. 1 (third and bottom), the reinitialization method (third) yields worse predictions than the concatenation method (bottom), especially after the replacement indicated by the red vertical lines, e.g. after time 300. Although the reinitialization method learns the initial parameters, $h_{\mathrm{init}}^{(\mathrm{p})}, h_{\mathrm{init}}^{(\mathrm{r})}, c_{\mathrm{init}}^{(\mathrm{p})}, c_{\mathrm{init}}^{(\mathrm{r})}$, its LSTM always attempts to forecast the degradation process similarly, thereby degrading performance.

4. Conclusion

We considered the degradation recovery process and examined feeding binary maintenance intervention records into LSTM. Experiments using synthetic data demonstrated the following: a binary representation is learnable for forecasting the process involving jumps through maintenance intervention. Furthermore, the concatenation method is a promising way to input such binary time series data into LSTM. This finding is useful for the development of a deep model for forecasting using data represented by binary time series.

Acknowledgments

This study was supported by JSPS KAKENHI Grant Number JP19K12039.

References

[1] K.L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," Mathematical Problems in Engineering, vol.2015, Article ID 793161, 2015.

[2] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to rul prediction," Mechanical Systems and Signal Processing, vol.104, pp.799–834, 2018.

[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol.9, no.8, pp.1735–1780, 1997.

[4] K. Kosukegawa, Y. Mori, H. Suyari, and K. Kawamoto, "Spatiotemporal forecasting of vertical track alignment with exogenous factors," Sci. Rep., vol.13, 2023.

[5] B. Castanier, C. Bérenguer, and A. Grall, "A sequential condition-based repair/replacement policy with non-periodic inspections for a system subject to continuous wear," Appl. Stochastic Models Bus. Ind., vol.19, no.4, pp.327–347, 2003.

[6] C. Park and W. Padgett, "Accelerated degradation models for failure based on geometric brownian motion and gamma processes," Lifetime Data Anal., vol.11, no.4, pp.511–527, 2005.

[7] D.P. Kingma and J.L. Ba, "Adam: A method for stochastic optimization," Proc. International Conference on Learning Representations (ICLR), 2015.

[8] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.2623–2631, 2019.

Appendix: Degradation Recovery Process

We define a degradation recovery process for structural health as follows. During the degradation process, we use the geometric Brownian motion (GBM) [6], which is a stochastic process given by a stochastic differential equation as $dx_t = \mu x_t dt + \sigma x_t dW_t$, where W_t is the Brownian motion and μ, σ are drift and volatility parameters, respectively. Using Ito's formula, the solution of GBM is given as $x_t = x_0 \exp(\tau t - \sigma W_t)$, where $\tau = \mu - \sigma^2/2$. Using the solution, we define the health degradation process as follows:

$$x_t^* = x_0 \exp\left(-\tau(t-s) - \sigma W_{t-s}\right) \quad \text{(A· 1)}$$

where x_0 denotes the initial health state, t is the current time and $s(< t)$ denotes the time of the last maintenance intervention, where τ denotes the degradation rate. The higher the τ, the faster the degradation. In addition, Gaussian noise $\varepsilon_t \sim N(0, \sigma^2)$ is added to x_t^* as $x_t = x_t^* + \varepsilon_t$.

During the recovery process, we sampled x_0 in Eq. (A· 1) from the normal distribution $N(1, 0.03^2)$. The recovery process occurs when either partial repair or replacement is performed. Partial repairs and replacements had varying effects on the degradation rate τ. The partial repair increases the degradation rate as $\tau' = \tau + \tau_d + \varepsilon_\tau$, τ_d is the incremental rate and $\varepsilon_\tau \sim N(0, 0.05^2)$. This increment increases τ each time a partial repair is performed, resulting in faster degradation. In contrast, replacement resets the increased degradation rate as $\tau' \sim N(1.0, 0.05^2)$. The choice of partial repair or replacement and its timing were determined as follows. If $x_t < L_d$ and $\tau > 2.5$, the replacement is performed immediately because the structure is severely degraded and in danger. If $x_t < L_d$ and $\tau < 2.5$, we perform partial repair because the degradation rate is not very high. Conversely, if $x_t > L_m$, we perform neither partial repair nor replacement. This becomes slightly more complicated when $L_d \leq x_t \leq L_m$. In this case, the structure is degraded but is not in severe danger. Therefore, we do not always perform maintenance interventions; instead, we decide whether to perform them probabilistically. We selected partial repair and replacement, each with a probability of 0.5.