Reinforcement learning in maintenance of civil infrastructures (ICML 2019)

Shiyin Wei [1] **Xiaowei Jin** [1] **Yuequan Bao** [1] **Hui Li** [1]

Abstract

Life-cycle management of aged civil infrastructures is an issue of worldwide concern. The process of sequential decision making on structural maintenance is usually considered as a Markov Decision Process (MDP) where Markov property holds in the structural condition transition due to the deterioration and maintenance. However, policy-making for large MDPs for maintenance of complex realistic infrastructures has long been a challenging problem due to the high-dimensions. Thus we introduce a deep reinforcement learning(DRL) framework to make this available, and a deep Q network implemented by CNN is employed to approximate the state-action value in the high-dimensional state-action space. A maintenance task of a cable-stayed bridge is designed and used to verify the efficiency of the proposed approach. The results show that the DRL is effective and efficient at the policy-making for maintenance tasks of complex civil infrastructures with high-dimensional state-action space.

1. Introduction

Civil infrastructure experiences performance degradation in serving due to environmental deterioratio or hazards. The structural condition assessment and maintenance processes have been the main concern throughout the world (Bao et al., 2019; Li & Ou, 2016).

Condition-based maintenance (CBM) policies using the inspected or predicted structural conditions (Li et al., 2018; Van Noortwijk, 2009; Wei et al., 2017) were studied. The CBM policies usually consist of a reward model (Frangopol et al., 2017), a condition predictive model and a core policy-decision model. The reward model measures the costs for different maintenance actions. The condition predictive model (Frangopol, 2011; Frangopol et al., 2004) measures the effectiveness of a maintenance action (including no repair) by predicting the associated consequences. The core policy-decision model is geared towards making rational, cost-effective, maintenance decisions based on the reward and condition information. Markov Decision Process (MDP) is frequently employed as the policy-decision model by asset management agencies in practice. For example, the core methodology behind the bridge management system (BMS) software packages PONTIS and BRIGIT are both MDP-based (Hawk, 1995; Mirzaei et al., 2014; Thompson et al., 1998).

However, the optimation is expensive and even impossible for problems with large state or action spaces using the traditional methods (such dynamic programming, and linear programming), thus the MDP-based BMSs are far to be put into realistic applications(Liu & Madanat, 2015; Madanat & Ibrahim, 1995; Papakonstantinou & Shinozuka, 2014). Thus, we propose a DRL framework for the automated policy-making of bridge maintenance actions, which is presented in Fig.1. Take the bridge maintenance as an example, the bridge maintenance system (BMS) corresponds to the agent, the bridge corresponds to the environment, whose structural condition will deteriorate due to natural environmental deterioration, hazards, and maintenance actions and enhance owing to maintenance action (which is the corresponding action output by the agent), the structural condition together with the serving year correspond to the state, and the weighted financial costs of maintenance and collapse risk corresponds to the reward. Duel Deep Q Network (Duel DQN) (Mnih et al., 2015; Wang et al., 2015) is employed as the optimation algorithm.

A key feature of the approach is summarized as: we establish the DRL framework for optimization of component-level maintenance policy, which is universal for maintenance tasks of various structures with little change in the network architecture. Although we employ a hand-craft maintenance environment and a model-free algorithm, the DRL framework can be easily transferred to a model-based case.

2. Preliminaries

As illustrated in Fig.1, we consider a standard reinforcement learning (RL) setup consisting of an agent and an

*Equal contribution [1]School of Civil Engineering, Harbin Institute of Technology, Harbin, China. Correspondence to: Shiyin Wei <shiyin.wei@ualberta.ca>.

Reinforcement Learning for Real Life (RL4RealLife) Workshop in the 36^{th} *International Conference on Machine Learning*, Long Beach, California, USA, 2019.

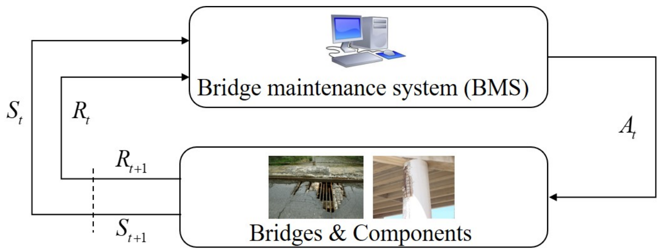

Figure 1. The structural maintenance task in RL perspective.

environment which the agent interacting with in discrete timesteps, and the timestep is set as a year. The DRL-based maintenance policy-making is based on MDP model, $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is the structural state space (i.e., the discrete structural rate set according to the inspection manual, such as very good, good, fair, poor, urgent and critical). The $\mathcal{A} = \{a_1, a_2, \cdots, a_m\}$ is the possible maintenance action space (i.e., the predefined maintenance actions with m types of action levels, such as no maintenance, minor-maintenance, major-maintenance, and replace). The $\mathcal{P}$ is the state transition probability matrix that indicating structural state transition probability from S_t at year t to S_{t+1} at year $t+1$ when performing the maintenance action A_t at year t. When the action A_t is 'no repair,'the structural state will be degraded by the probability due to the erosion of natural environments or hazards. When the action A_t is a maintenance of different levels (minor-maintenance, major-maintenance, or replace), the structural states will be enhanced to different levels by the probability due to the maintenance. The policy is defined as a conditional probability of action under the given state, $\pi\left(A_t|S_t\right) = P\left(A_t|S_t\right)$, and $\gamma \in [0, 1]$ is the discount factor considering the long-term rewards. The goal of DRL-based maintenance policy-making is to learn the maintenance policy that maximizes the total reward during the entire lifespan of the structure. All the dynamical and reward model of the maintenance task are predefined based on the inspection and maintenance manual and the cost criterion, while the state transition model is obtained from practical experiences or a computation model of the structure.

A sequence of bridge states, maintenance actions, and rewards that depict the entire history of the RL task following a certain policy π is denoted as an episode as given by $S_1, A_1, R_1, S_2, \cdots, R_{T-1}, S_T, A_T, R_T \sim \pi$. The optimal policy should consider all the sequences during the entire lifespan of the bridge. Therefore, return $G_t$that balances the short and long-term rewards (where T is the lifespan of the bridge) and the state-action value function $Q_\pi\left(S_t, A_t\right)$ that calculates the expectation are both introduced as:

$$\begin{gathered} G_t = R_{t+1} + \gamma R_{t+2} + \cdots = \textstyle\sum_{k=0}^{T-t-1} \gamma^k R_{t+k+1} \\ Q_\pi\left(S_t, A_t\right) = E_\pi\left[G_t | S_t, A_t\right] \end{gathered} \tag{1}$$

A DNN -structured Q network $\hat{Q}_{\mathbf{w}}\left(s, a\right) \approx Q_\pi\left(s, a\right)$ is employed to approximate the Q value in this study due to the powerful nonlinear representation and mapping capacity (Mnih et al., 2015; Schmidhuber, 2015). A Duel DQN architecture(Wang et al., 2015) is employed and trained.

3. Examples and results

We conduct an example application on a hand-craft maintenance environment of a long-span cable-stayed bridge with a main span of 648 m in mainland China (Chen et al., 2016; Li et al., 2018) to investigate the application of the proposed approach in complex structures. The investigated bridge is a two-tower cable-stayed bridge with 168 cables, 89 box-girder sections, two bridge towers, and four bridge piers, as shown in Fig.2, giving a total of 263 components.

3.1. Problem Formulation

STATES

Structural conditions for all components are rated from 0 to 5 according to the severity of the material defects and the percentage loss of the component cross-section and surface area along length; the rank 0 is 'very good' and 5 is 'critical.' In addition, decisions can be different for the same structural condition during a different lifecycle year, and the serving year (age) of the bridge is another important factor in the maintenance policy decision-making process. Therefore, the state $\mathcal{S}$ in the DRL framework stacks the components' structural condition rates and serving year and forms a vector. A trick here is to encode the structural condition of each component to the one-hot encoding vector of length 6, and the serving year (with the scale of 100) is encoded to the binary form as a vector of 7 so that the states are normalized. It is then expanded to a vector of length 1,600 by zero padding with a length of 15, and reshaped to a 2-D vector of 40×40. Thus, the size of the state space is $|\mathcal{A}| = 100 \times 6^{263}$.

MAINTENANCE ACTIONS

Actions are generally divided into four discrete levels according to the maintenance effect and costs (Papakonstantinou & Shinozuka, 2014):no-repair (code 0), minor-maintenance (code 1), major-maintenance (code 2), and replacement (code 3). Therefore, each component has one option among the four maintenance actions. Thus, the size of the action space is $|\mathcal{A}| = 4^{263}$.

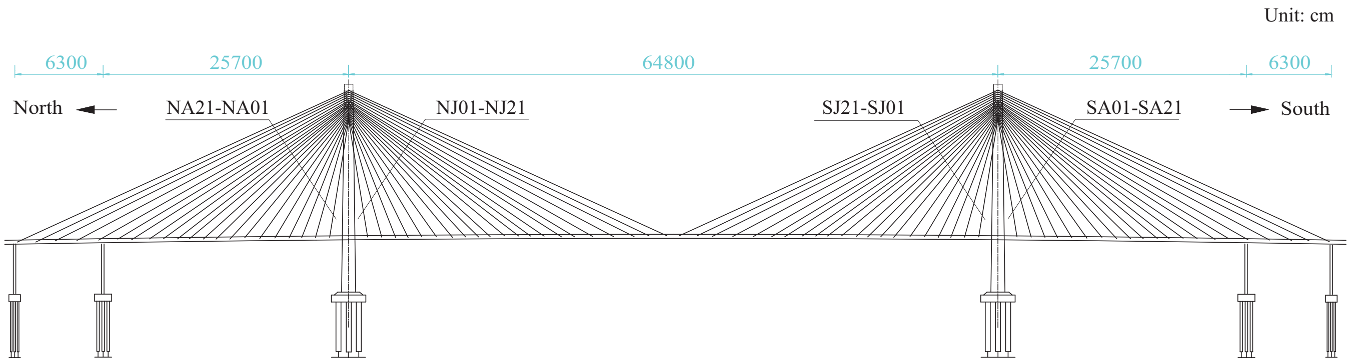

Figure 2. The example cable-stayed bridge(Li et al., 2018)

$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}0.92 & 0.08 & 0 & 0 & 0 & 0\\ 0 & 0.91 & 0.09 & 0 & 0 & 0\\ 0 & 0 & 0.91 & 0.09 & 0 & 0\\ 0 & 0 & 0 & 0.90 & 0.10 & 0\\ 0 & 0 & 0 & 0 & 0.90 & 0.10\\ 0 & 0 & 0 & 0 & 0 & 1.00\end{bmatrix}\end{array}$$

(a) Stay-cables

$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}0.85 & 0.15 & 0 & 0 & 0 & 0\\ 0 & 0.84 & 0.16 & 0 & 0 & 0\\ 0 & 0 & 0.81 & 0.19 & 0 & 0\\ 0 & 0 & 0 & 0.78 & 0.22 & 0\\ 0 & 0 & 0 & 0 & 0.75 & 0.25\\ 0 & 0 & 0 & 0 & 0 & 1.00\end{bmatrix}\end{array}$$

(b) Box girders

$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}0.96 & 0.04 & 0 & 0 & 0 & 0\\ 0 & 0.94 & 0.06 & 0 & 0 & 0\\ 0 & 0 & 0.92 & 0.08 & 0 & 0\\ 0 & 0 & 0 & 0.90 & 0.10 & 0\\ 0 & 0 & 0 & 0 & 0.89 & 0.11\\ 0 & 0 & 0 & 0 & 0 & 1.00\end{bmatrix}\end{array}$$

(c) Towers

$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}0.95 & 0.05 & 0 & 0 & 0 & 0\\ 0 & 0.93 & 0.07 & 0 & 0 & 0\\ 0 & 0 & 0.91 & 0.09 & 0 & 0\\ 0 & 0 & 0 & 0.90 & 0.10 & 0\\ 0 & 0 & 0 & 0 & 0.88 & 0.12\\ 0 & 0 & 0 & 0 & 0 & 1.00\end{bmatrix}\end{array}$$

(d) Piers

Figure 3. State transition probability matrices under moderate and no maintenance for cable-stayed bridge

$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}1.0 & 0 & 0 & 0 & 0 & 0\\ 0.80 & 0.20 & 0 & 0 & 0 & 0\\ 0.40 & 0.50 & 0.10 & 0 & 0 & 0\\ 0.20 & 0.30 & 0.40 & 0.10 & 0 & 0\\ 0.10 & 0.20 & 0.30 & 0.32 & 0.08 & 0\\ 0.05 & 0.10 & 0.15 & 0.32 & 0.35 & 0.03\end{bmatrix}\end{array}$$

(a) Minor-maintenance

$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}1.0 & 0 & 0 & 0 & 0 & 0\\ 1.0 & 0 & 0 & 0 & 0 & 0\\ 0.90 & 0.08 & 0.02 & 0 & 0 & 0\\ 0.70 & 0.15 & 0.10 & 0.05 & 0 & 0\\ 0.60 & 0.18 & 0.12 & 0.07 & 0.03 & 0\\ 0.50 & 0.25 & 0.15 & 0.06 & 0.03 & 0.01\end{bmatrix}\end{array}$$

(b) Major-maintenance

$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\end{bmatrix}\end{array}$$

(c) Repair

Figure 4. State transition probability matrices under different level of maintenance for cable-stayed bridge

STATE TRANSITION

The effect of maintenance actions and natural/hazard deterioration are assumed to be dependent on the structural conditions and the maintenance actions, and is determined by the maintenance level via the state transition matrices. Therefore, the worse the condition is, the worse the maintenance effect is under the same maintenance action, and the higher the level the maintenance action is, the better the maintenance effect is under the same structural condition. A replacement maintenance transforms a component from any condition to a condition of 0. The hand-craft state transition probability matrices, as shown in Fig.3 and Fig.4.

REWARDS

Rewards are defined as a combination of the financial maintenance costs and structural risks in Eq.2. The negative value is taken to be consistent with the literal meaning of 'reward.' Rewards are the user specified goals for maintaining a bridge, and the optimal policy leads to maintenance actions that maximize the total rewards. Reference values can be made based on an elaborate statistic on the costs provided in http://sv08data.dot.ca.gov/contractcost/, or specified by the manager. However, the exact value is not important in this study, so the costs are specified by the authors. The maintenance cost of a component c is assumed to be dependent on the state s and the maintenance actions a as illustrated in Eq.2.

$$\begin{aligned} R\left(c,s,a\right) = cost_{total}(c) \times rate_{condition}(s) \times rate_{action}(a) \\ + cost_{total}(c) \times rate_{risk}(s) \end{aligned} \tag{2}$$

where $cost_{total}(c), rate_{condition}(c), rate_{action}(a)$ are the cost rates with dependencies on the structural components c, state of structural components s, and level of maintenance actions a respectively; the $rate_{risk}(s)$ is the structural risk rate related to bridge states, and $cost_{total}(c) \times rate_{risk}(s)$ measures the risk by probabilistic financial costs. The units for the economic cost are set to 1 since the relative values are more important than the absolute ones in this study, and the summation of costs of $cost_{total}(c)$ over all components is normalised to 1.

ARCHITECTURE

The DRL architecture is shown in textcolorblueFig.5 with CNN-structured feature learning procedure and fully-connected Q-learning procedure. The state S_t is treated as the input and the state-action value $Q(S_a, a)$ is approximated by the outputs of the network. The ϵ-greedy policy is employed to sample the maintenance action in the training step. The input stacks the one-hot encoded structural condition with a length of 1,578 (263×6) and the binary-encoded serving year with a length of 7, then is expanded to a vector of length 1,600 by zero padding with a length of 15. Finally, it is reshaped to $40 \times 40 \times 1$. The hyper parameters are set as $\gamma = 0.95, \alpha = 0.0001$, the capacity of the experience buffer is 10^5, and the training batch size is 10^3. The network is trained on a 64-bit Ubuntu desktop enabled with a GTX-1080Ti GPU device.

3.2. Results

The performance of the DRL is shown in Fig.**??**, and the DRL policy converges to the policy with the lowest cost after 100,000 training steps (training time of approximately 3 days) from a random initialized policy.

Figure 7 compares the normalized costs for different maintenance policies (DRL, time-based, and condition-based maintenance policies) over 1,000 Monte-Carlo simulations of the bridge from the beginning t=1 to terminal t=100, the 'Time-5' indicates the time-based policies that suggest minor-maintenance actions on all deck components every 5 years (similarly for the Time-10, Time-15, and Time-20 policies). It can be seen that the DRL finds the optimal results among all the policies, where 'Condition-1' is the best of the hand-crafted policies in the cable-stayed bridge maintenance environment. The maintenance performances for different policies are shown by the lifecycle condition and action distribution in Fig.6, the DRL policy and Condition-1 policy also lead to similar lifecycle condition distributions which are are mainly kept to be in either condition 1 or condition 0, and the DRL policy tends to have less frequent maintenance actions because of the balance of risk expectation and costs of maintenance. These results imply that the DRL is effective in finding the optimal policy for different maintenance tasks.

4. Related Work

Some notable approaches in the field of structural maintenance policies are mainly focused on the physical deterioration models(Guo et al., 2015), known as the model predictive control methods(Frangopol et al., 2017; Edirisinghe et al., 2015; Zhang et al., 2017). The MDP based policy-making researches are mostly based on DP or LP algorithms that limits in small state and action spaces. Recently, Papakonstantinou et.al have tried to combine DRL method and the structural health monitoring technique in structural maintenance (which is a classic partially observed MDP problem) (Andriotis & Papakonstantinou, 2018). In summary, the application of DRL in civil engineering and maintenance policies has just started, and lots of work including problem formulation, and history-data collection has to be studied.

5. Conclusions

This paper proposes a general deep reinforced learning (DRL) framework for structural maintenance. An example for a long-span cable-stayed bridge are given, and the following conclusions are made:

- This paper proposes a DRL framework for a general solution to the high-dimensional component-level maintenance policy decision for civil infrastructures.
- The DRL framework provides a general method for structures with different complexities (with different numbers of components) with little change in the network architecture.
- The model-free DRL framework is available to directly learn from real historical data or an environment model. Physical or structural concerns (hazard events like earthquakes) should be put into the task environment if it is necessary. It is important to have reasonable deterioration models and cost criteria to obtain a rational maintenance policy through the learning process.

6. Acknowledgements

This study was financially supported by the NSFC (Grant No. U1711265 and 51638007).

References

Andriotis, C. and Papakonstantinou, K. Managing engineering systems with large state and action spaces through deep reinforcement learning. *arXiv preprint arXiv:1811.02052*, 2018.

Bao, Y., Chen, Z., Wei, S., Xu, Y., Tang, Z., and Li, H. The state of the art of data science and engineering in structural health monitoring. *Engineering*, 2019.

Chen, Z., Bao, Y., and Li, H. The random field model of the spatial distribution of heavy vehicle loads on long-span bridges. In *Health Monitoring of Structural and*

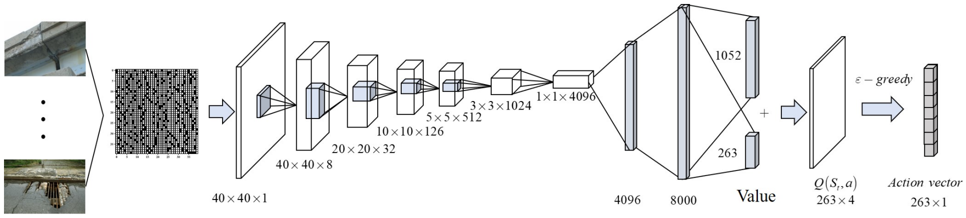

Figure 5. The DRL architecture for the maintenance task of the cable-stayed bridge

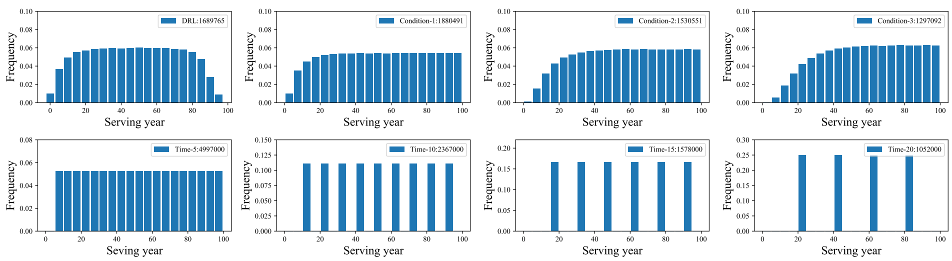

Figure 6. Comparison of different maintenance policies for 1,000 simulations: condition distribution

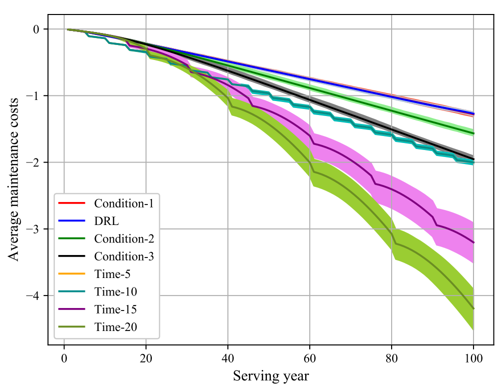

Figure 7. Average maintenance costs of the different policies by serving year

Biological Systems 2016, volume 9805, pp. 98052Q. International Society for Optics and Photonics, 2016.

Edirisinghe, R., Setunge, S., and Zhang, G. Markov modelbased building deterioration prediction and iso factor analysis for building management. *Journal of Management in Engineering*, 31(6):04015009, 2015.

Frangopol, D. M. Life-cycle performance, management, and optimisation of structural systems under uncertainty: accomplishments and challenges 1. *Structure and Infrastructure Engineering*, 7(6):389–413, 2011.

Frangopol, D. M., Kallen, M.-J., and Noortwijk, J. M. v. Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. *Progress in structural engineering and Materials*, 6(4):197–212, 2004.

Frangopol, D. M., Dong, Y., and Sabatino, S. Bridge lifecycle performance and cost: analysis, prediction, optimisation and decision-making. *Structure and Infrastructure Engineering*, 13(10):1239–1257, 2017.

Guo, A., Yuan, W., Lan, C., Guan, X., and Li, H. Timedependent seismic demand and fragility of deteriorating bridges for their residual service life. *Bulletin of Earthquake Engineering*, 13(8):2389–2409, 2015.

Hawk, H. Bridgit deterioration models. *Transportation Research Record*, 1490:19–22, 1995.

Li, H. and Ou, J. The state of the art in structural health monitoring of cable-stayed bridges. *Journal of Civil Structural Health Monitoring*, 6(1):43–67, 2016.

Li, S., Wei, S., Bao, Y., and Li, H. Condition assessment of cables by pattern recognition of vehicle-induced cable tension ratio. *Engineering Structures*, 155:1–15, 2018.

Liu, H. and Madanat, S. Adaptive optimisation methods in system-level bridge management. *Structure and Infrastructure Engineering*, 11(7):884–896, 2015.

Madanat, S. and Ibrahim, W. H. W. Poisson regression models of infrastructure transition probabilities. *Journal of Transportation Engineering*, 121(3):267–272, 1995.

Mirzaei, Z., Adey, B. T., Klatter, L., and Thompson, P. The iabmas bridge management committee overview of existing bridge management systems, 2014.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529, 2015.

Papakonstantinou, K. and Shinozuka, M. Optimum inspection and maintenance policies for corroded structures using partially observable markov decision processes and stochastic, physically based models. *Probabilistic Engineering Mechanics*, 37:93–108, 2014.

Schmidhuber, J. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

Thompson, P. D., Small, E. P., Johnson, M., and Marshall, A. R. The pontis bridge management system. *Structural engineering international*, 8(4):303–308, 1998.

Van Noortwijk, J. A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1):2–21, 2009.

Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., and De Freitas, N. Dueling network architectures for deep reinforcement learning. *arXiv preprint arXiv:1511.06581*, 2015.

Wei, S., Zhang, Z., Li, S., and Li, H. Strain features and condition assessment of orthotropic steel deck cable-supported bridges subjected to vehicle loads by using dense fbg strain sensors. *Smart Materials and Structures*, 26(10):104007, 2017.

Zhang, Y., Kim, C.-W., and Tee, K. F. Maintenance management of offshore structures using markov process model with random transition probabilities. *Structure and Infrastructure Engineering*, 13(8):1068–1080, 2017.