Network-Scale Maintenance Planning for Infrastructures Using Reinforcement Learning (Formulation, Part 2)

Zachary Hamida Postdoc
James-A. Goulet Professeur

Polytechnique Montréal, Canada
Département des génies civil, géologique et des mines

August 6, 2021

Funding:
Transportation Ministry of Quebec (MTQ)

Outline

Recap & Definitions

Reinforcement Learning - Context

Problem Formulation

Recap & Definitions

Visual Inspections (VI): Network-scale monitoring technique

Source: Google images

Recap & Definitions

Visual Inspections (VI): Network-scale monitoring technique

Network:

Source: Google images

Recap & Definitions

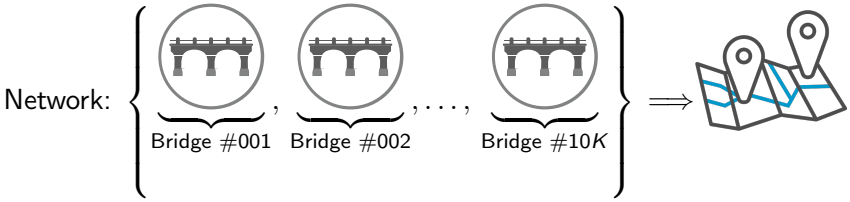
Visual Inspections (VI): Network-scale monitoring technique

Network: $\left\{ \underbrace{\text{  }}_{\text{Bridge \#001}}, \underbrace{\text{  }}_{\text{Bridge \#002}}, \ldots, \underbrace{\text{  }}_{\text{Bridge \#10}K} \right\}$

Source: Google images

Recap & Definitions

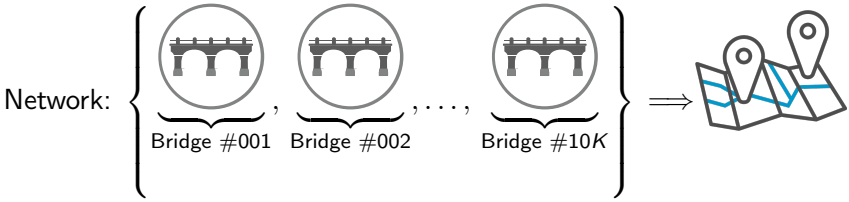
Visual Inspections (VI): Network-scale monitoring technique



Source: Google images

Recap & Definitions

Visual Inspections (VI): Network-scale monitoring technique

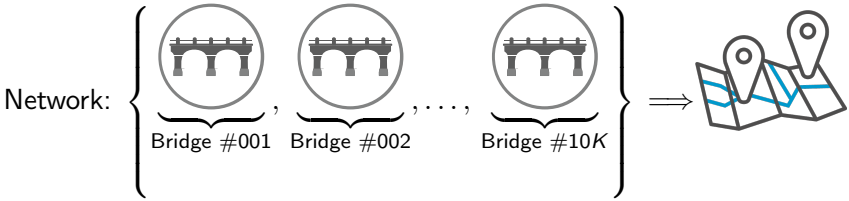


Bridge: {type, material, traffic, location,...}

Source: Google images

Recap & Definitions

Visual Inspections (VI): Network-scale monitoring technique

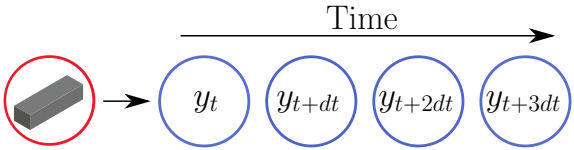


Bridge: {type, material, traffic, location,...}

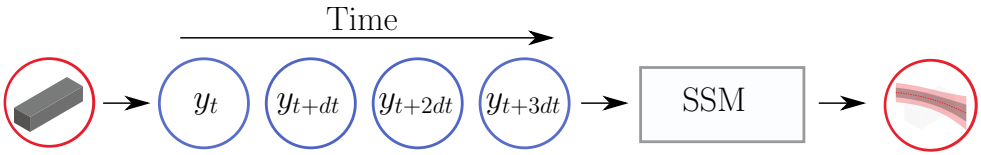
Source: Google images

Deterioration Framework

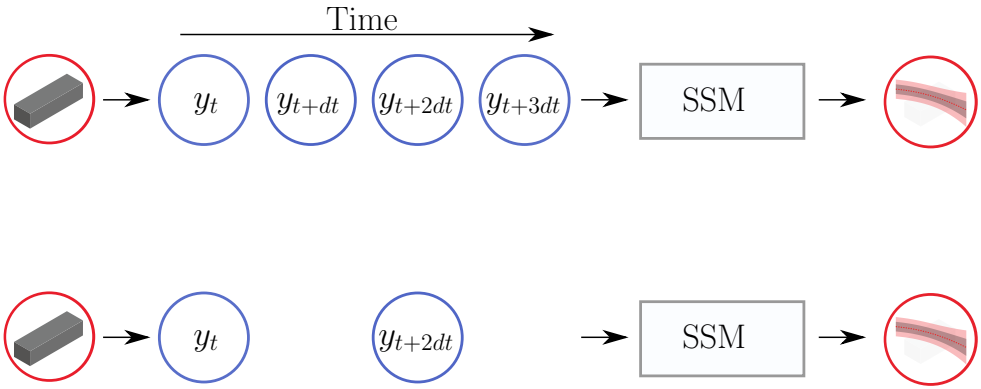
Deterioration Framework



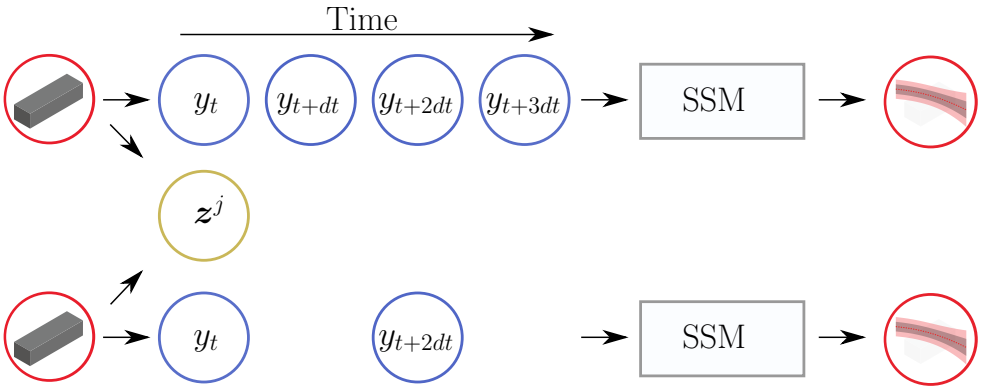
Deterioration Framework



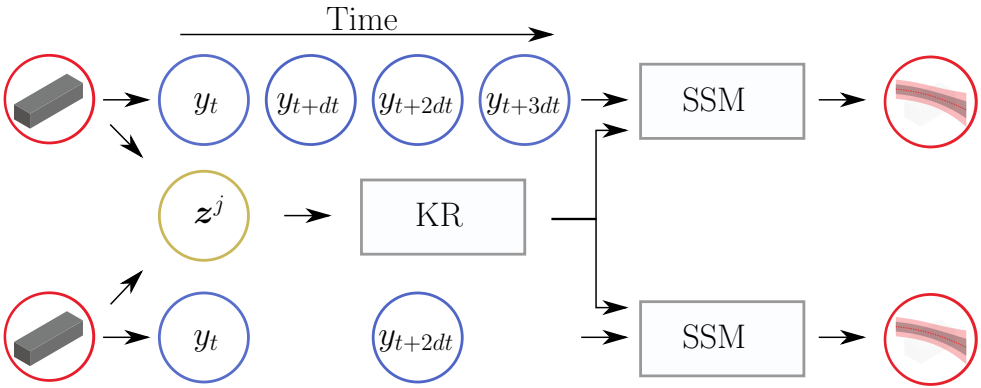
Deterioration Framework



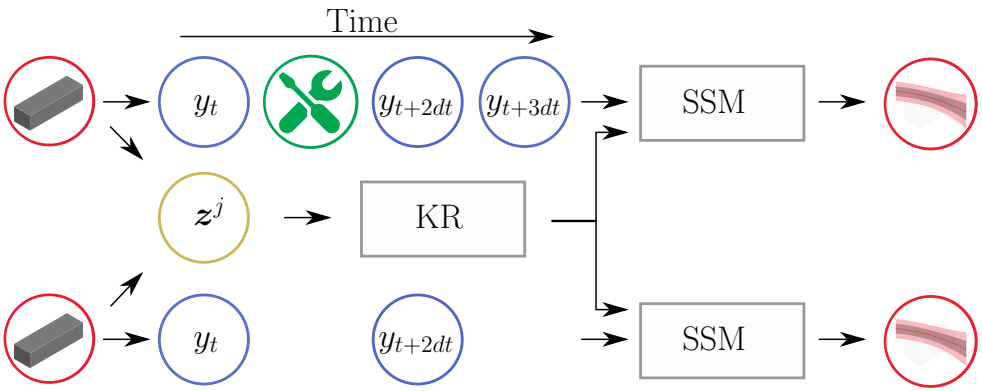
Deterioration Framework



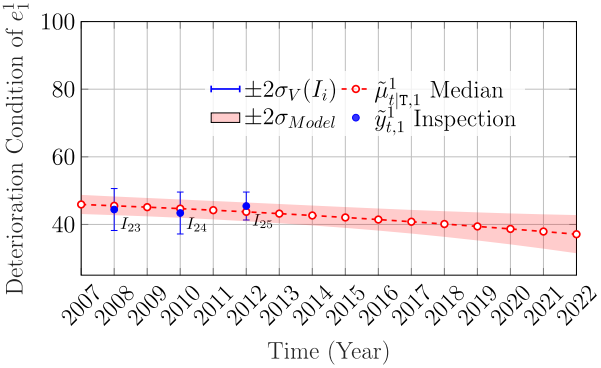
Deterioration Framework



Deterioration Framework

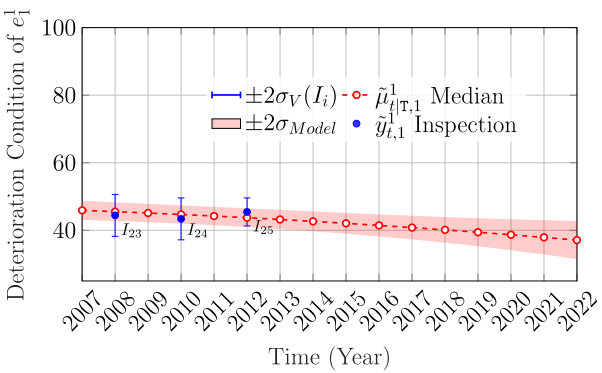


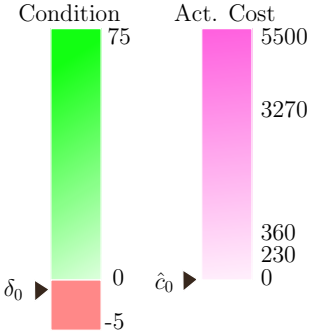
Element-Level Analyses



Source: MTQ, Manual of Inspection

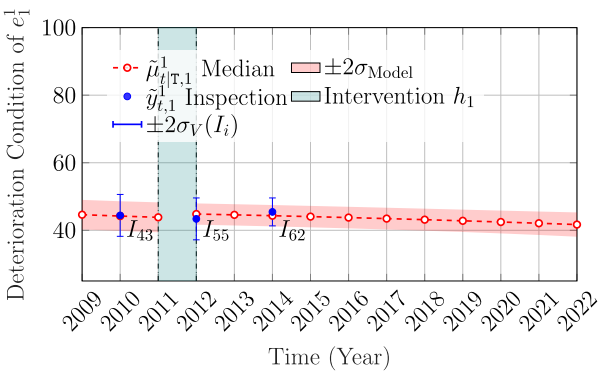
Element-Level Analyses

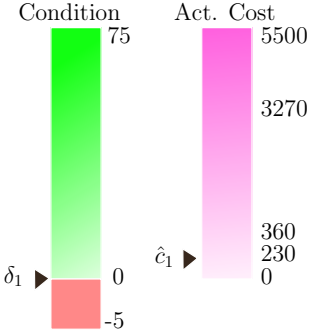




Source: MTQ, Manual of Inspection

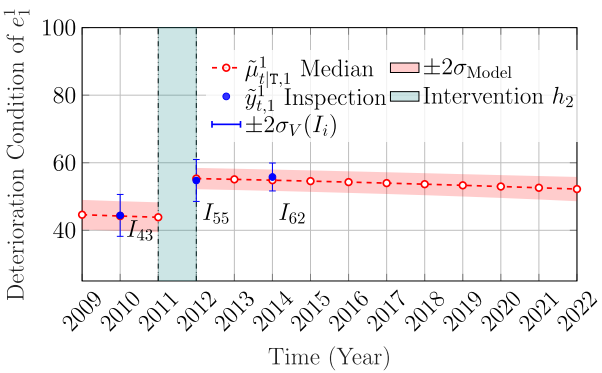
Element-Level Analyses

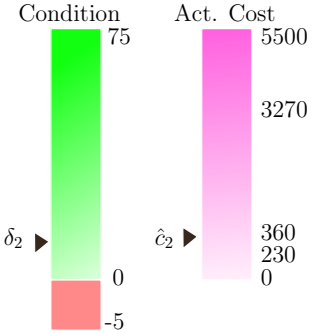




Source: MTQ, Manual of Inspection

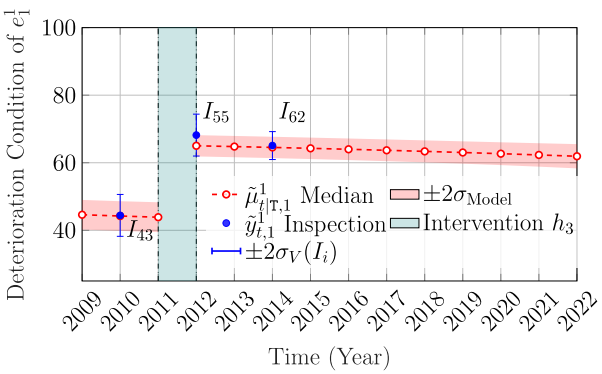
Element-Level Analyses

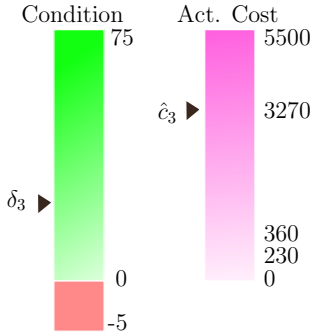




Source: MTQ, Manual of Inspection

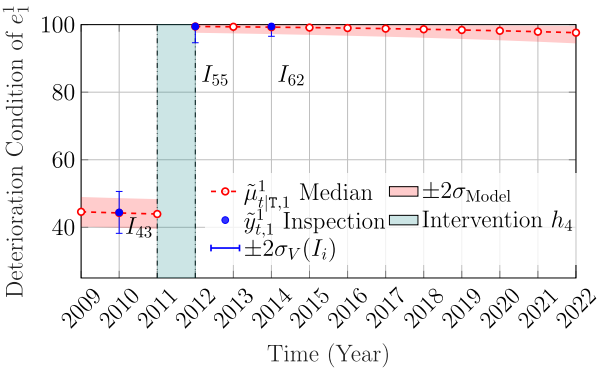
Element-Level Analyses

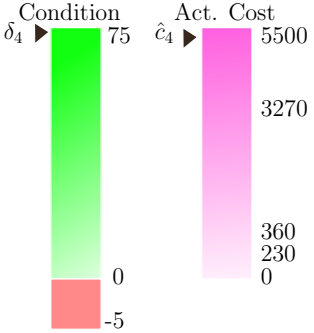




Source: MTQ, Manual of Inspection

Element-Level Analyses

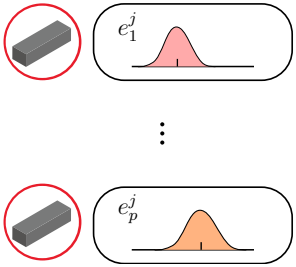




Source: MTQ, Manual of Inspection

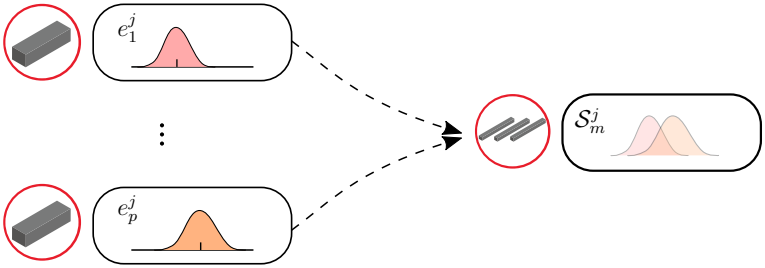
System-Level Analyses

At any time t:



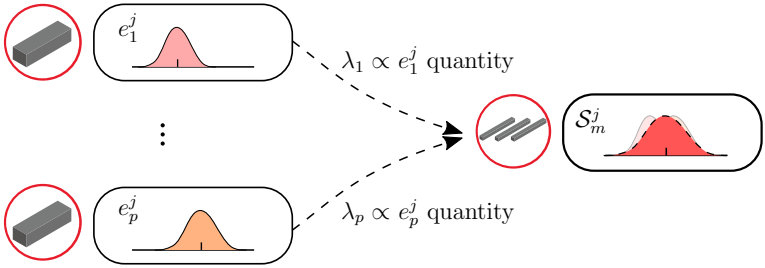
System-Level Analyses

At any time t:

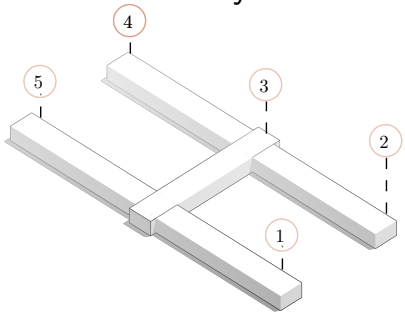


System-Level Analyses

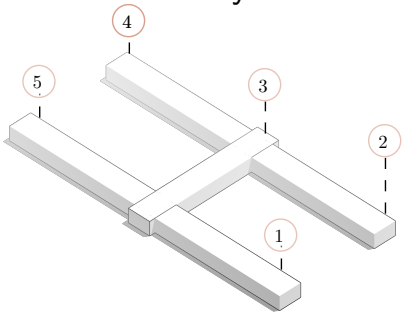
At any time t:

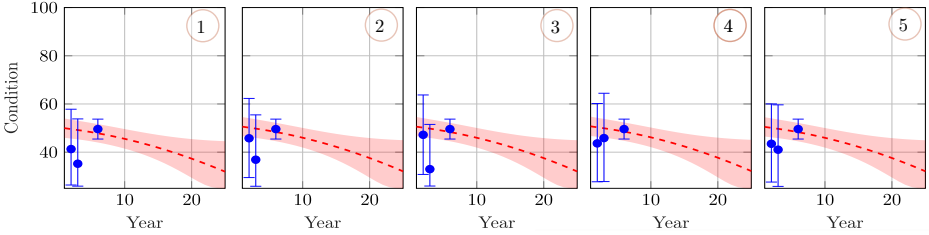


System-Level Analyses

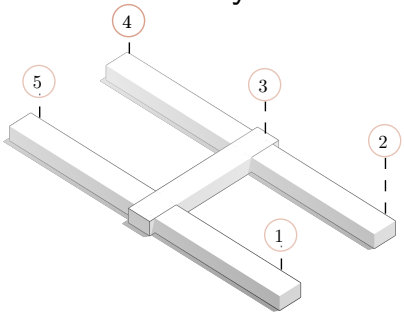


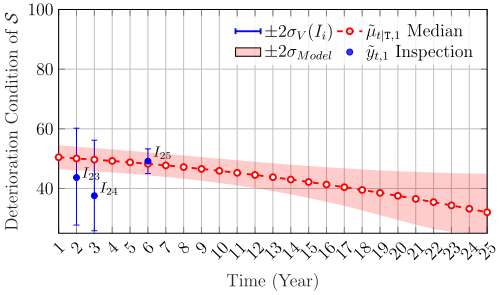
System-Level Analyses

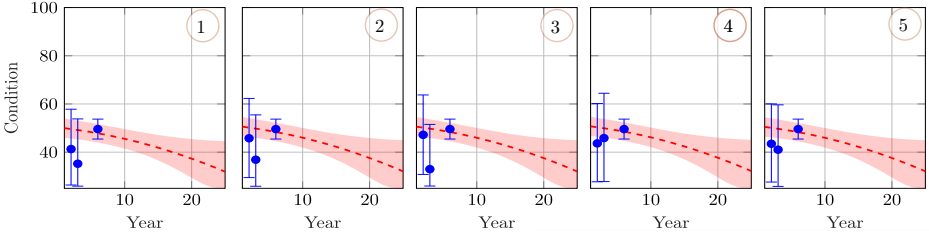




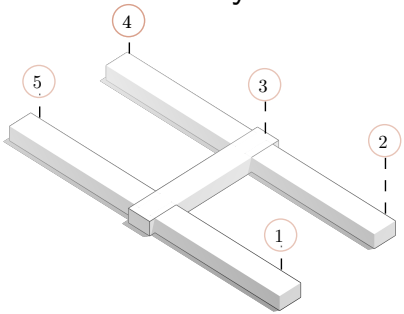
System-Level Analyses

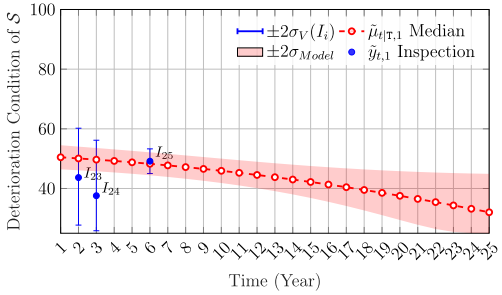


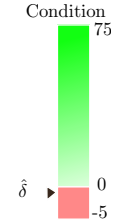


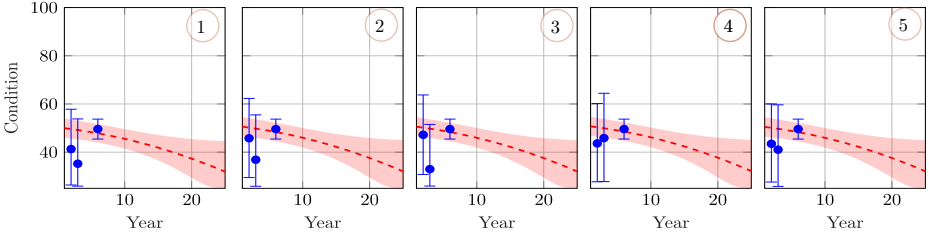


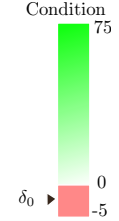
System-Level Analyses



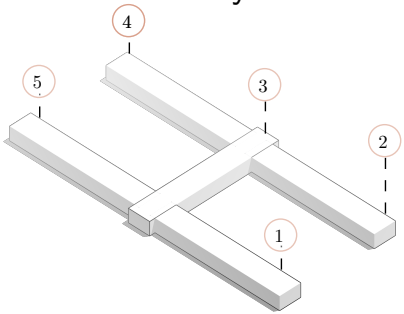


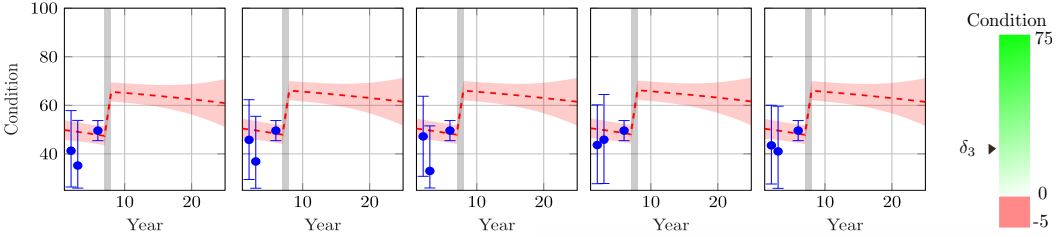




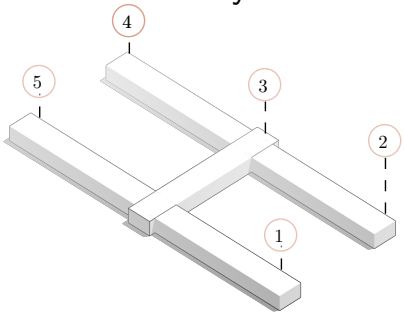


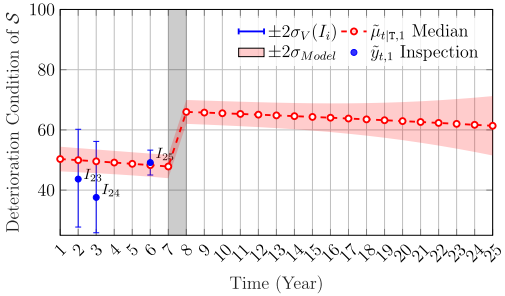
System-Level Analyses

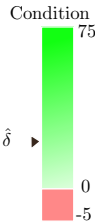


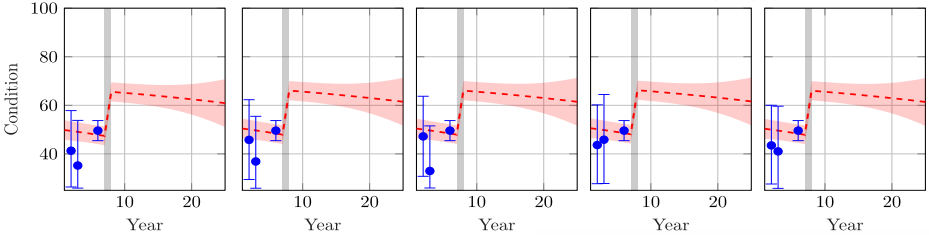


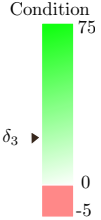
System-Level Analyses











Research Objective

Find a policy for maintenance:

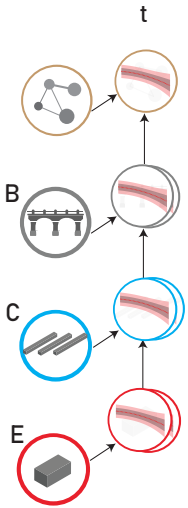
$$\pi\left(\text{  }, \text{  }, \text{  }, \text{  }, \text{  }\right)$$

.

Research Objective

Find a policy for maintenance:

$$\pi \left(\text{[network]}, \text{B}, \text{C}, \text{E}, \$ \right)$$

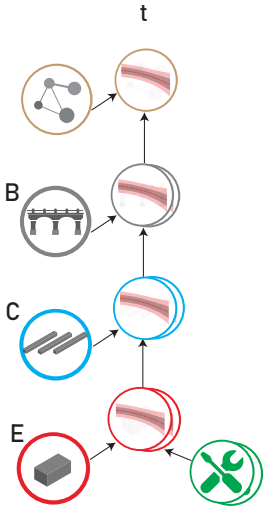


.

Research Objective

Find a policy for maintenance:

π (, B, C, E, $)

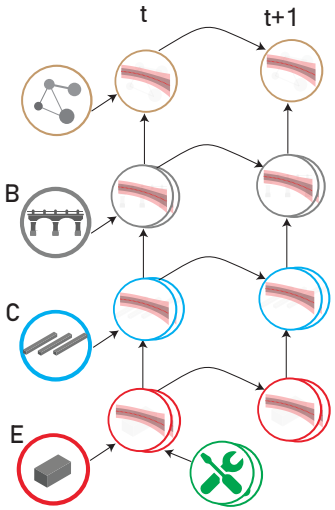


.

Research Objective

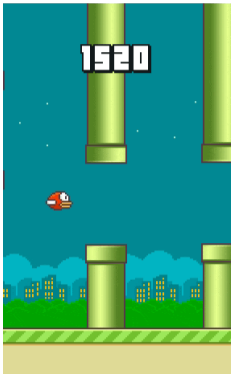
Find a policy for maintenance:

$$\pi \left(\text{network}, \text{B}, \text{C}, \text{E}, \$ \right)$$



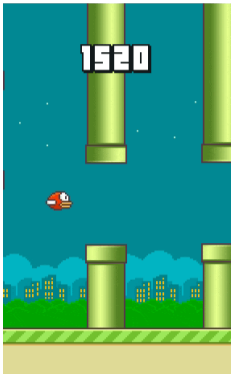
.

Reinforcement Learning



Source: Google Images

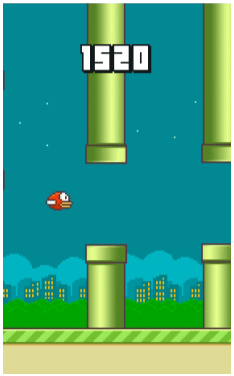
Reinforcement Learning



- Agent: decision maker.

Source: Google Images

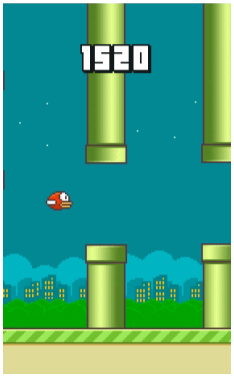
Reinforcement Learning



- ▷ Agent: decision maker.
- ▷ Environment: e.g., video game.

Source: Google Images

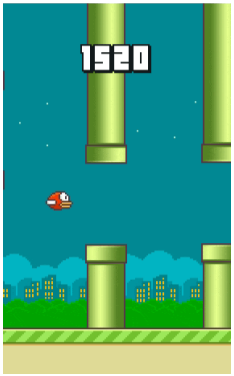
Reinforcement Learning



- ▷ Agent: decision maker.
- ▷ Environment: e.g., video game.
- ▷ Actions: e.g., key press or movement.

Source: Google Images

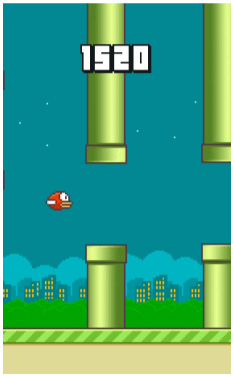
Reinforcement Learning



- ▷ Agent: decision maker.
- ▷ Environment: e.g., video game.
- ▷ Actions: e.g., key press or movement.
- ▷ Rewards: e.g., achieving the goal.

Source: Google Images

Reinforcement Learning



- ▷ Agent: decision maker.
- ▷ Environment: e.g., video game.
- ▷ Actions: e.g., key press or movement.
- ▷ Rewards: e.g., achieving the goal.

Learn a policy that maximizes the total expected discounted rewards.

Source: Google Images

Environment

Employing the deterioration model as an environment requires:

Environment

Employing the deterioration model as an environment requires:

- ▷ Adjustments for the actions' exploration:

Environment

Employing the deterioration model as an environment requires:

- ▷ Adjustments for the actions' exploration:
  - ✓ Account for the boundaries of the feasible condition.

Environment

Employing the deterioration model as an environment requires:

- ▷ Adjustments for the actions' exploration:
  - ✓ Account for the boundaries of the feasible condition.
  - ✓ Account for repeating actions.

Environment

Employing the deterioration model as an environment requires:

- ▷ Adjustments for the actions' exploration:
  - ✓ Account for the boundaries of the feasible condition.
  - ✓ Account for repeating actions.
- ▷ Defining measures of success:

Environment

Employing the deterioration model as an environment requires:

- ▷ Adjustments for the actions' exploration:
  - ✓ Account for the boundaries of the feasible condition.
  - ✓ Account for repeating actions.
- ▷ Defining measures of success:
  - ✓ Compatibility (Elements and Category).

Environment

Employing the deterioration model as an environment requires:

- ▷ Adjustments for the actions' exploration:
  - ✓ Account for the boundaries of the feasible condition.
  - ✓ Account for repeating actions.
- ▷ Defining measures of success:
  - ✓ Compatibility (Elements and Category).
  - ✓ Criticality (Elements and Category).

Environment

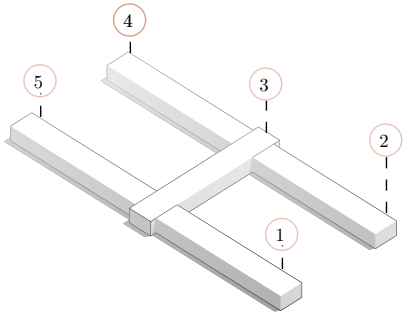
Employing the deterioration model as an environment requires:

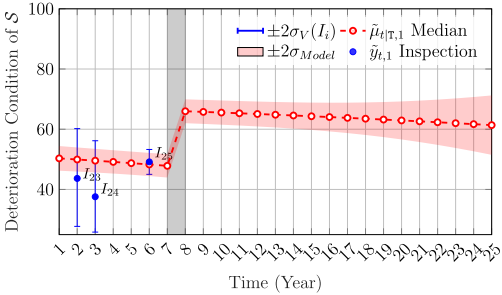
- ▷ Adjustments for the actions' exploration:
  - ✓ Account for the boundaries of the feasible condition.
  - ✓ Account for repeating actions.
- ▷ Defining measures of success:
  - ✓ Compatibility (Elements and Category).
  - ✓ Criticality (Elements and Category).
  - ✓ Variability (Elements).

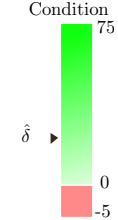
Environment

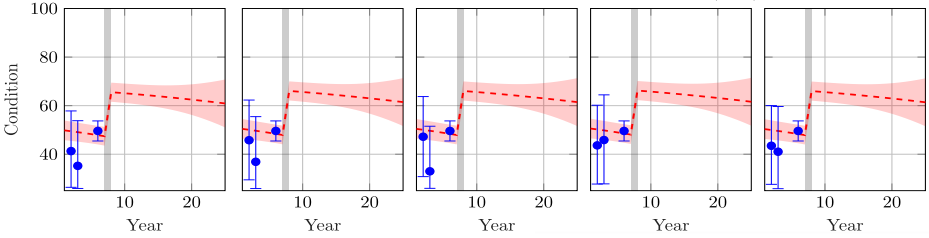
Employing the deterioration model as an environment requires:

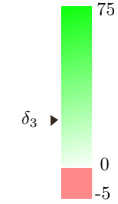
- ▷ Adjustments for the actions' exploration:
  - ✓ Account for the boundaries of the feasible condition.
  - ✓ Account for repeating actions.
- ▷ Defining measures of success:
  - ✓ Compatibility (Elements and Category).
  - ✓ Criticality (Elements and Category).
  - ✓ Variability (Elements).
  - ✓ Action frequency (Category).

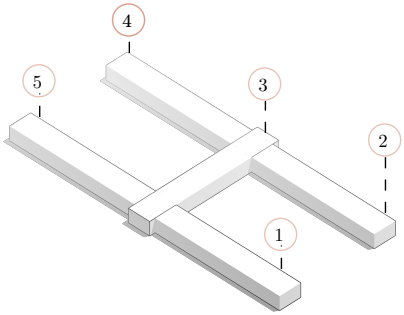

4
5
3
2
1

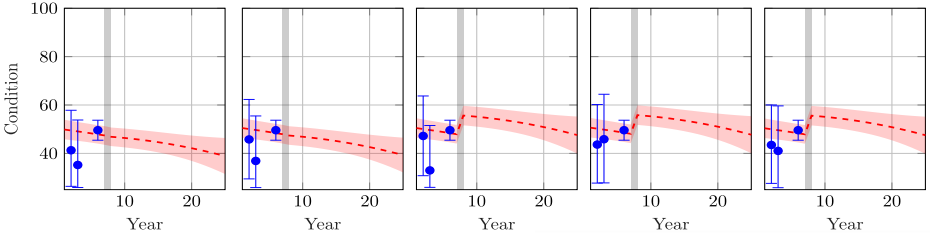

Deterioration Condition of $\mathcal{S}$
$\pm 2\sigma_V(I_i)$
$\tilde{\mu}_{t|\mathrm{T},1}$ Median
$\pm 2\sigma_{Model}$
$\tilde{y}_{t,1}$ Inspection
I_{23}
I_{24}
Time (Year)

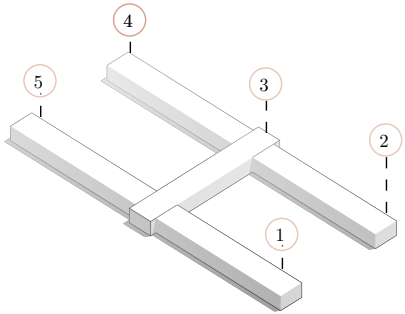

Condition
75
$\hat{\delta}$
0
-5

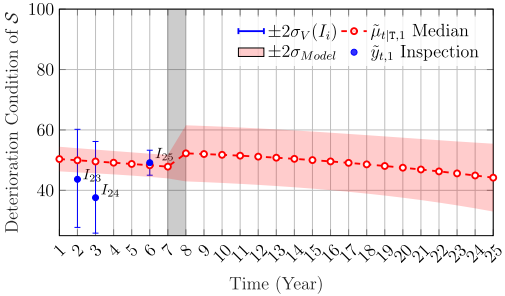

Condition
Year

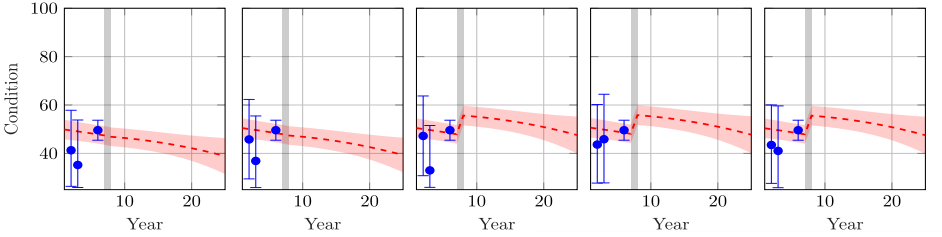

Condition
75
δ_3
0
-5

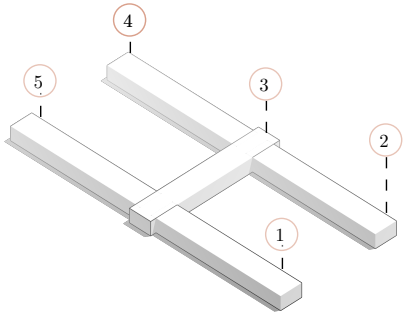

4
5
3
2
1

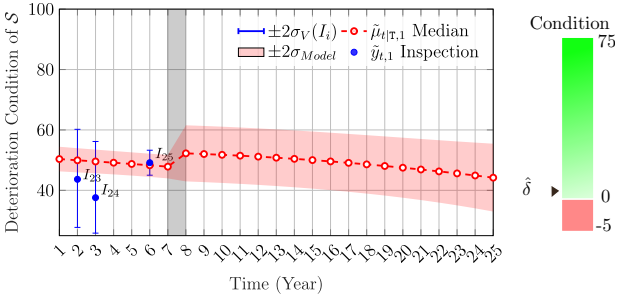

Condition
100
80
60
40
10
20
Year
Condition
75
0
-5
δ_2
δ_1

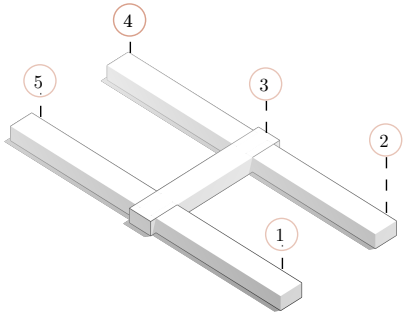

4
5
3
2
1

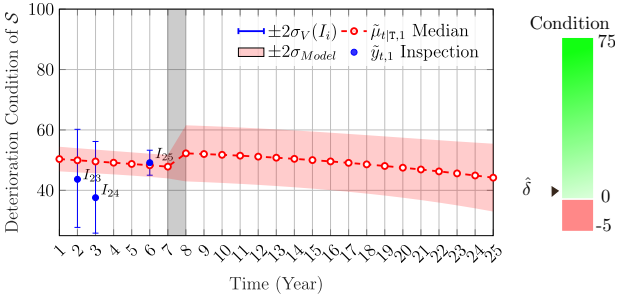

Deterioration Condition of $\mathcal{S}$
$\pm 2\sigma_V(I_i)$
$\pm 2\sigma_{Model}$
$\tilde{\mu}_{t|\mathbf{T},1}$ Median
$\tilde{y}_{t,1}$ Inspection
I_{23}
I_{24}
I_{25}
Time (Year)
Condition
$\hat{\delta}$
75
0
-5


Condition
Year
Condition
δ_2
δ_1
75
0
-5

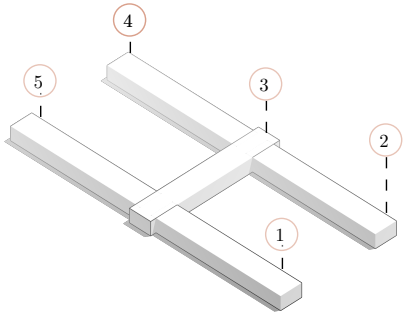

4
5
3
2
1

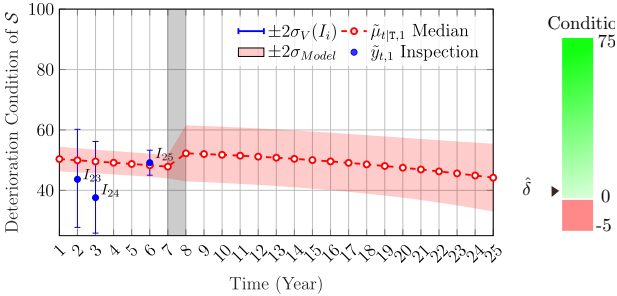

Deterioration Condition of $\mathcal{S}$
100
80
60
40
$\pm 2\sigma_V(I_i)$
$\pm 2\sigma_{Model}$
$\tilde{\mu}_{t|\mathbf{T},1}$ Median
$\tilde{y}_{t,1}$ Inspection
I_{25}
I_{23}
I_{24}
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Time (Year)
Condition
75
0
-5
$\hat{\delta}$



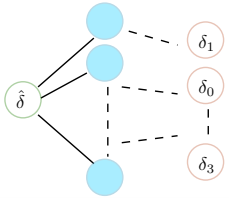


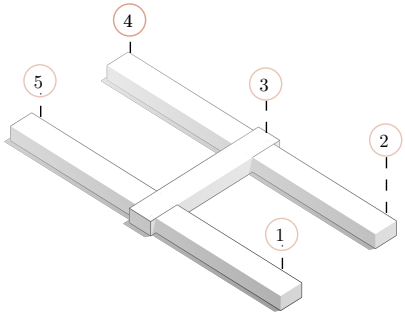
$$L = \|\mathbb{E}[\boldsymbol{\delta}] - \hat{\delta}\|$$

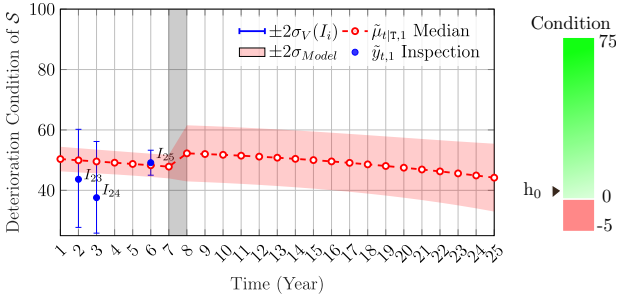




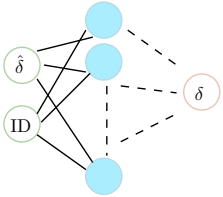
$$L = \|\mathbb{E}[\boldsymbol{\delta}] - \hat{\delta}\|$$

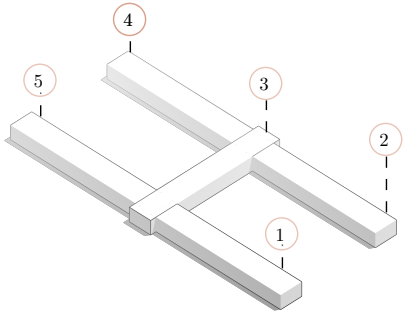


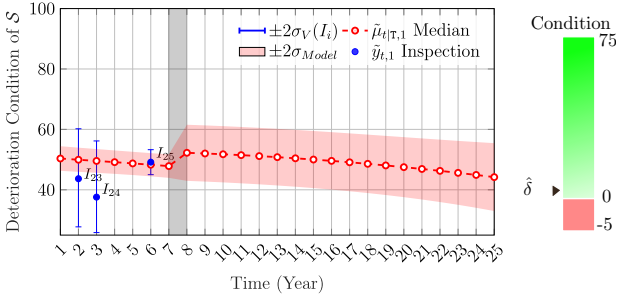




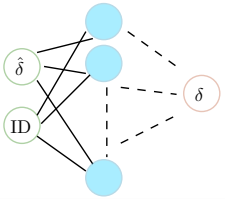
$$L = \|\frac{\delta_i}{\mathbf{E}} - \hat{\delta}\|$$







$$L = \|\tfrac{\delta_i}{\mathbf{E}} - \hat{\delta}\| \times \hat{c}_i$$



RL Sub-task (Local Policy π^e):

RL Sub-task (Local Policy π^e):

- ▷ state: $\hat{\delta}_i$, ID

RL Sub-task (Local Policy π^e):

- ▷ state: $\hat{\delta}_i$, ID
- ▷ actions: a_0, a_1, a_2, a_3, a_4 corresponding to the effects $\delta_0, \delta_1, \delta_2, \delta_3, \delta_4$

RL Sub-task (Local Policy π^e):

- ▷ state: $\hat{\delta}_i$, ID
- ▷ actions: a_0, a_1, a_2, a_3, a_4 corresponding to the effects $\delta_0, \delta_1, \delta_2, \delta_3, \delta_4$
- ▷ rewards: $\hat{r}_i = -\|\hat{\delta} - \frac{\delta_i}{\mathrm{E}}\| \times \hat{c}_i$.

RL Sub-task (Local Policy π^e):

- ▷ state: $\hat{\delta}_i$, ID
- ▷ actions: a_0, a_1, a_2, a_3, a_4 corresponding to the effects $\delta_0, \delta_1, \delta_2, \delta_3, \delta_4$
- ▷ rewards: $\hat{r}_i = -\|\hat{\delta} - \frac{\delta_i}{\mathrm{E}}\| \times \hat{c}_i$.
- ▷ transition model: $\hat{\delta}_{i+1} = \hat{\delta}_i - \frac{\delta_i}{\mathrm{E}}$.

RL Sub-task (Local Policy π^e):

- ▷ state: $\hat{\delta}_i$, ID
- ▷ actions: a_0, a_1, a_2, a_3, a_4 corresponding to the effects $\delta_0, \delta_1, \delta_2, \delta_3, \delta_4$
- ▷ rewards: $\hat{r}_i = -\|\hat{\delta} - \frac{\delta_i}{\mathbf{E}}\| \times \hat{c}_i$.
- ▷ transition model: $\hat{\delta}_{i+1} = \hat{\delta}_i - \frac{\delta_i}{\mathbf{E}}$.

Use Q-learning to find the policy π^e for the RL problem.

RL Sub-task (Local Policy π^e):

- ▷ state: $\hat{\delta}_i$, ID
- ▷ actions: a_0, a_1, a_2, a_3, a_4 corresponding to the effects $\delta_0, \delta_1, \delta_2, \delta_3, \delta_4$
- ▷ rewards: $\hat{r}_i = -\|\hat{\delta} - \frac{\delta_i}{\mathrm{E}}\| \times \hat{c}_i$.
- ▷ transition model: $\hat{\delta}_{i+1} = \hat{\delta}_i - \frac{\delta_i}{\mathrm{E}}$.

Use Q-learning to find the policy π^e for the RL problem.

But, what about the maintenance policy π ?

RL Main Task (Global Policy π):

RL Main Task (Global Policy π):

- ▷ state: $\mu_{t|\mathrm{T}}, \dot{\mu}_{t|\mathrm{T}}$

RL Main Task (Global Policy π):

- ▷ state: $\mu_{t|\mathrm{T}}, \dot{\mu}_{t|\mathrm{T}}$
- ▷ actions: $\hat{\delta}_t$ (continuous action)

RL Main Task (Global Policy π):

- ▷ state: $\mu_{t|\mathrm{T}}, \dot{\mu}_{t|\mathrm{T}}$
- ▷ actions: $\hat{\delta}_t$ (continuous action)
- ▷ rewards: $r_t = \sum_i^{\mathrm{E}} \hat{c}_i$.

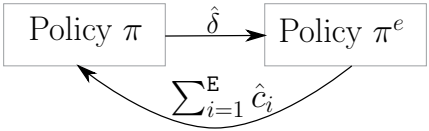
RL Main Task (Global Policy π):

- ▷ state: $\mu_{t|\mathrm{T}}, \dot{\mu}_{t|\mathrm{T}}$
- ▷ actions: $\hat{\delta}_t$ (continuous action)
- ▷ rewards: $r_t = \sum_i^{\mathrm{E}} \hat{c}_i$.

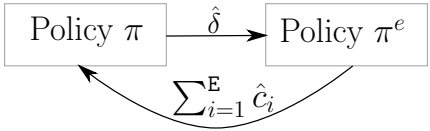
Use Q-learning (with policy gradient) to solve the RL problem.

Simultaneous Training for π^e and π (HRL):

Simultaneous Training for π^e and π (HRL):

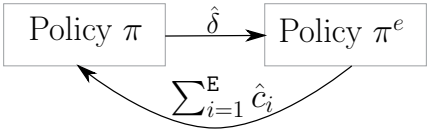


Simultaneous Training for π^e and π (HRL):



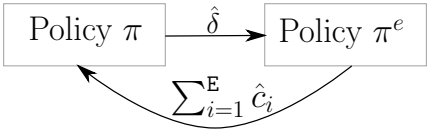
Problem: $\|\mathbb{E}[\delta] - \hat{\delta}\| \gg 0$

Simultaneous Training for π^e and π (HRL):



Problem: $\|\mathbb{E}[\delta] - \hat{\delta}\| \gg 0 \rightarrow \sum_{i=1}^{\mathbf{E}} \hat{c}_i$ and $\hat{\delta}$ do not correspond to each other.

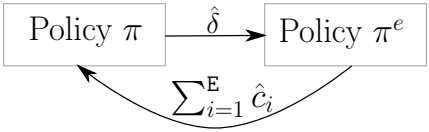
Simultaneous Training for π^e and π (HRL):



Problem: $\|\mathbb{E}[\delta] - \hat{\delta}\| \gg 0 \rightarrow \sum_{i=1}^{\mathbf{E}} \hat{c}_i$ and $\hat{\delta}$ do not correspond to each other.

Solution: replace $\hat{\delta}$ with $\bar{\delta}$ that maximizes the log probability $\log \pi^e(a_i|\bar{\delta})$ as in:

Simultaneous Training for π^e and π (HRL):

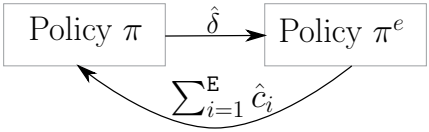


Problem: $\|\mathbb{E}[\delta] - \hat{\delta}\| \gg 0 \rightarrow \sum_{i=1}^{\mathbf{E}} \hat{c}_i$ and $\hat{\delta}$ do not correspond to each other.

Solution: replace $\hat{\delta}$ with $\bar{\delta}$ that maximizes the log probability $\log \pi^e(a_i|\bar{\delta})$ as in:

$$\log \pi^e(a_i|\bar{\delta}) \propto -\frac{1}{2} \sum_{i=1}^{\mathbf{E}} \|a_i - \pi^e(\bar{\delta}_i, \mathrm{ID})\|$$

Simultaneous Training for π^e and π (HRL):



Problem: $\|\mathbb{E}[\delta] - \hat{\delta}\| \gg 0 \rightarrow \sum_{i=1}^{\mathtt{E}} \hat{c}_i$ and $\hat{\delta}$ do not correspond to each other.

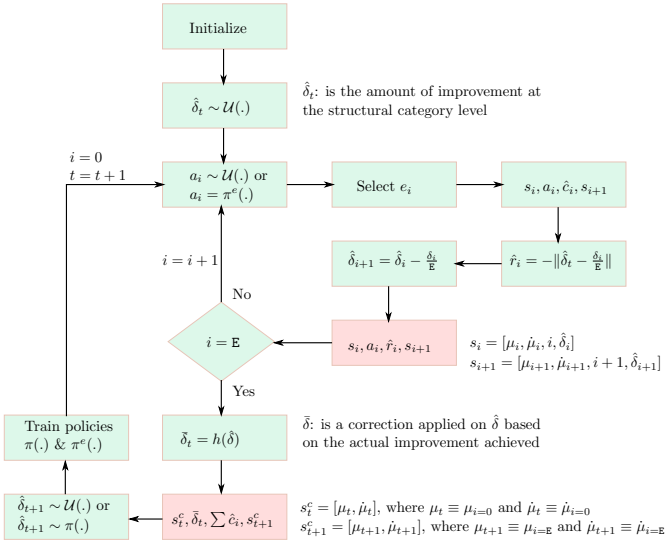
Solution: replace $\hat{\delta}$ with $\bar{\delta}$ that maximizes the log probability $\log \pi^e(a_i|\bar{\delta})$ as in:

$$\log \pi^e(a_i|\bar{\delta}) \propto -\frac{1}{2} \sum_{i=1}^{\mathtt{E}} \|a_i - \pi^e(\bar{\delta}_i, \mathrm{ID})\|$$

To maximize this quantity, candidate goals $\bar{\delta}$ are sampled based on the original goal $\hat{\delta}$.

Full HRL Framework

Full HRL Framework



Next Steps

- Improve environment design (e.g., repeating actions).

Next Steps

- ▷ Improve environment design (e.g., repeating actions).
- ▷ Design the reward functions.

Next Steps

- ▷ Improve environment design (e.g., repeating actions).
- ▷ Design the reward functions.
- ▷ Account for the budget in the framework.