How Does Data Freshness Affect Real-time Supervised Learning?

Md Kamran Chowdhury Shisher and Yin Sun

Dept. of ECE, Auburn University

Question 1: Does Learning Error Increase with Data Staleness?

Theoretical and Experimental Findings:

Case 1: If the feature and target **data sequence** is a **Markov** chain, then

- Learning errors (**Training error and Inference error**) **increase monotonically** as **feature ages**.

Case 2: If The feature and target data sequence is **far from Markovian** (due to **Communication delay**, **Response delay**, and **long-range dependence**), then

- Learning errors **may not increase monotonically** as **feature becomes stale**.

Age of information (AoI): **Time difference** between the **current time** and the **generation time** of the **freshest received feature**.

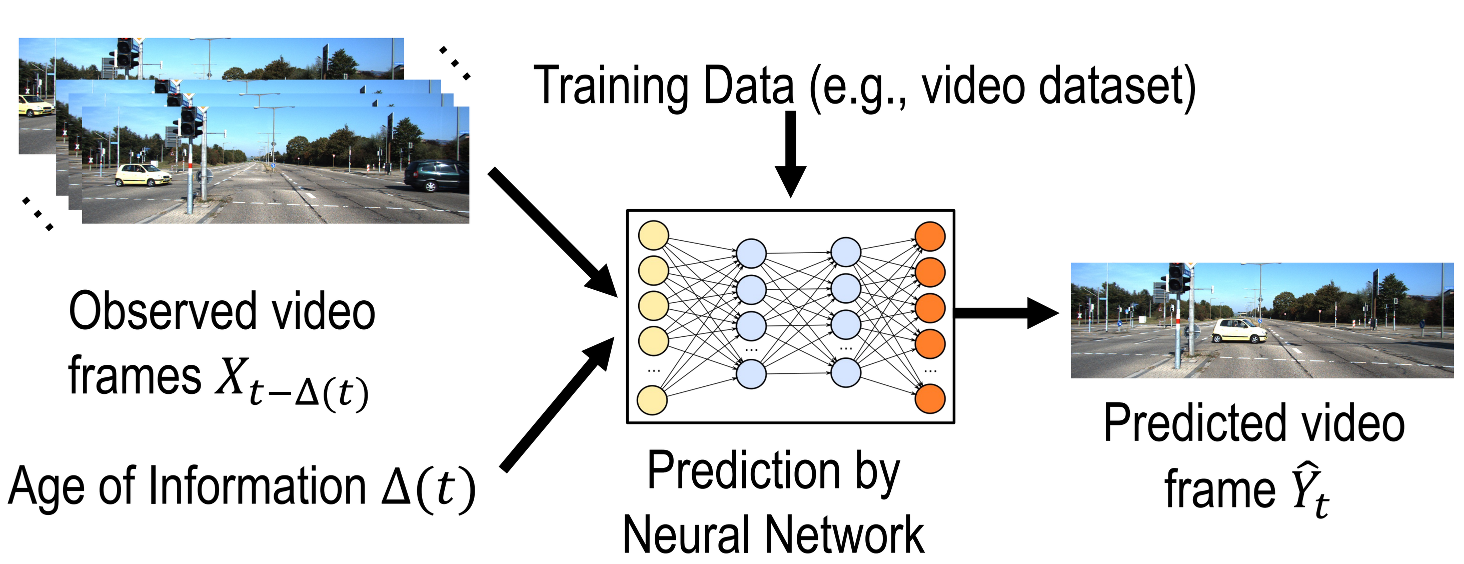
Theorem 1: The learning error is the **difference of two increasing functions** of AoI δ.

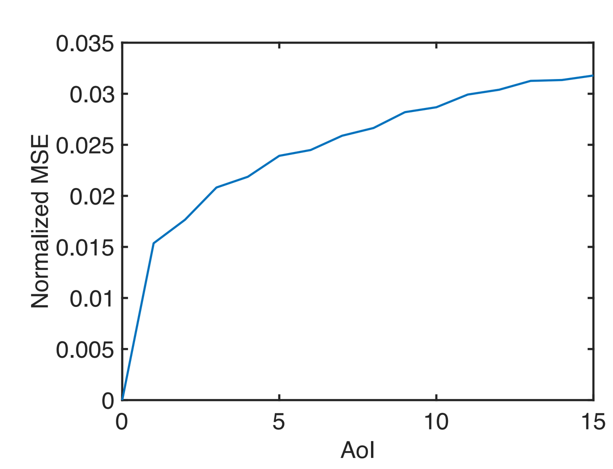
$$p(\delta) = g_1(\delta) - g_2(\delta).$$

If the data sequence is close to **Markovian**, function $g_2(\delta)$ is close to **0**.

Theorem 1 is proved by using **information theoretic** approach.

Inference Error increases with AoI in Video Prediction

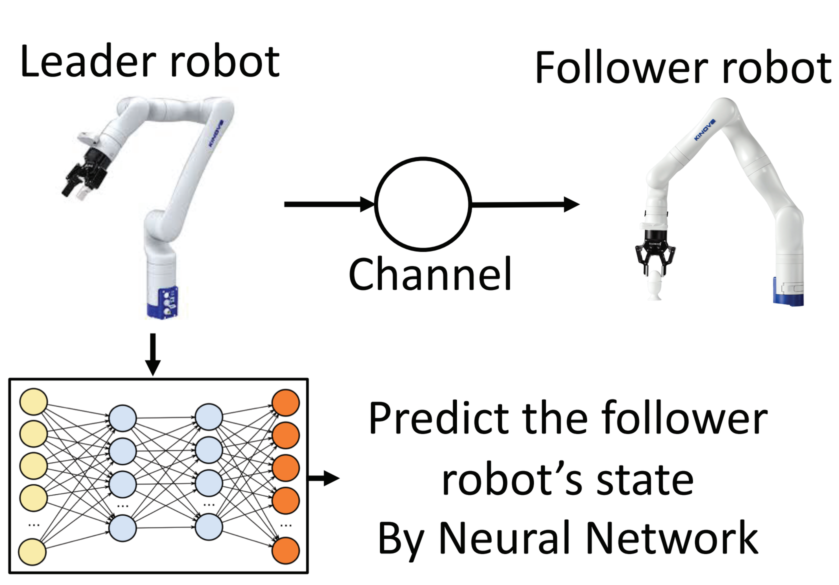




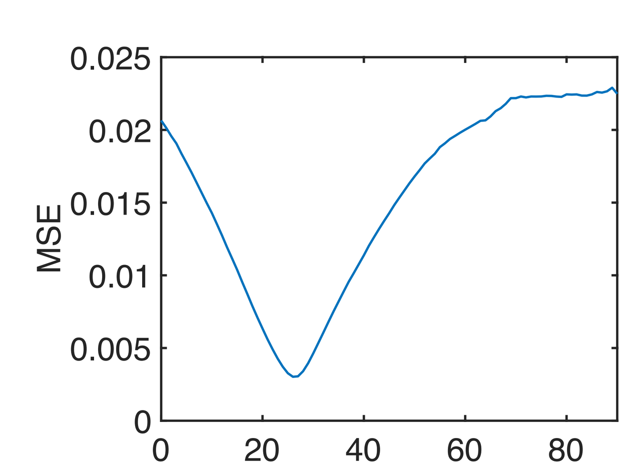
Inference Error vs. AoI

Video sequence is approximately a **Markov chain**.

Robot State Prediction in a Leader-follower Robotic System



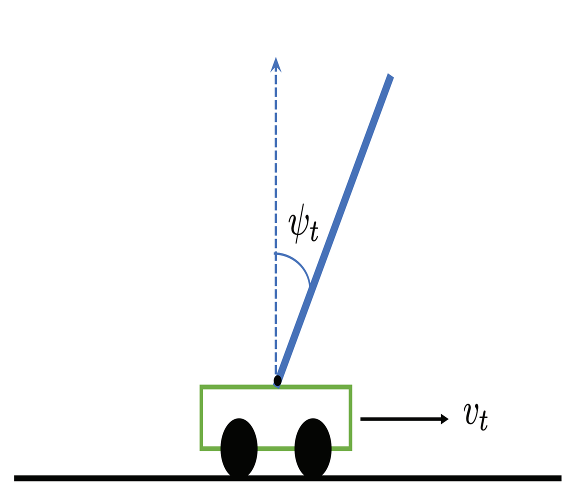
Prediction of Follower Robot



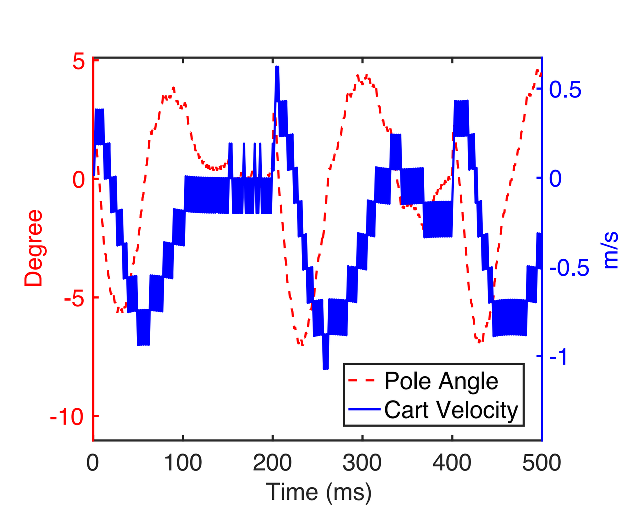
Inference Error vs. AoI

Inference Error decreases in the AoI ≤ 25 and increases when AoI ≥ 25.

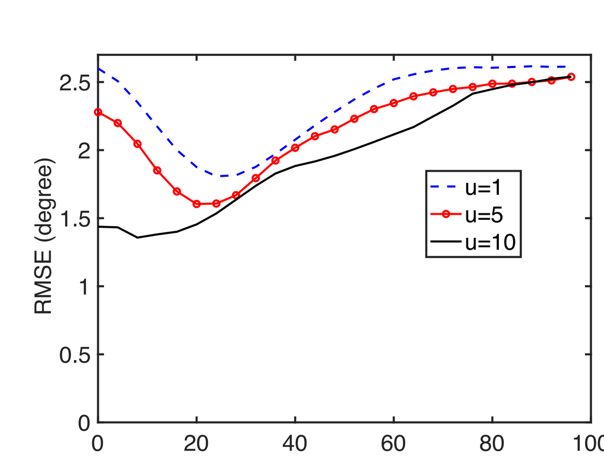
Actuator State Prediction under Response Delay



Cart pole Balancing



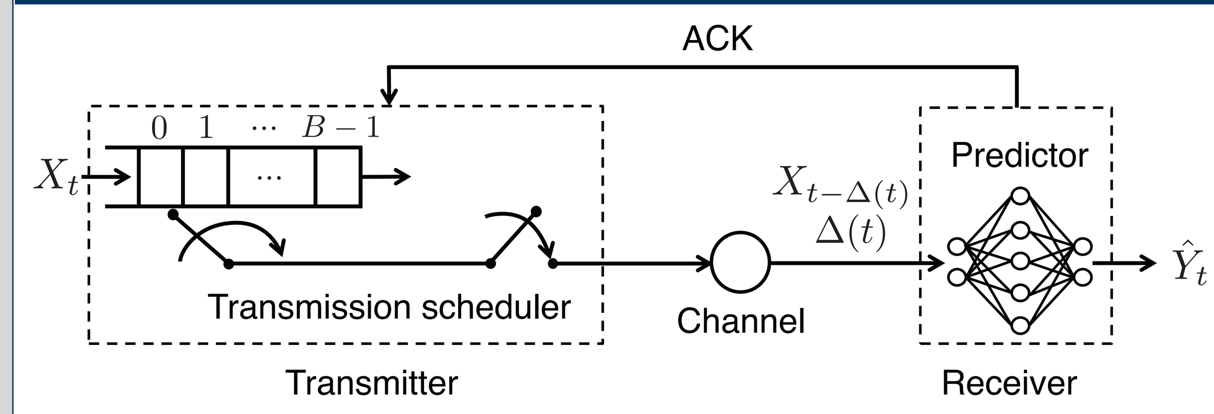
Data Traces



Inference Error vs. AoI

- Data traces show **25-30 ms** response delay.
- The data sequence is **non-Markovian**.
- The inference error becomes **increasing function** as the length u of feature $X_t = \{s_t, \ldots, s_{t-u+1}\}$ **increases**.

Question 2: Optimal Feature Transmission



Findings:

- Sending **old feature** could be **better**.
- Send another feature when Gittins index $\gamma(\Delta(t))$ exceeds a threshold.

Optimal Scheduling Policy

Theorem 2: The **optimal policy** is

$$S_{i+1} = \inf_{t \in \mathbb{Z}} \{ t \geq D_i : \gamma(\Delta(t)) \geq p_{\text{opt}} \},$$

where optimal objective value p_{opt} is a solution of the following **fixed point equation** of β :

$$\beta = \frac{\mathbb{E}\left[\sum_{t=D_i(\beta)}^{D_{I+1}(\beta)-1} p(\Delta(t))\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}$$

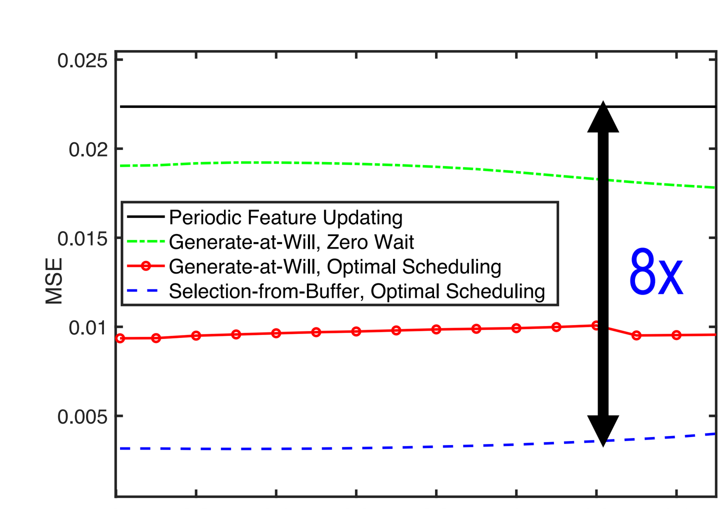
Low Complexity Algorithm

The $(i+1)$-th feature is sent at the earliest time slot t satisfying two conditions:

- The i-th feature has already been delivered by time slot t, i.e., $t \geq D_i$.
- The **Gittins index** $\gamma(\Delta(t))$ exceeds p_{opt}.
- The **optimal position of the buffer** is **constant** for all i and is solved analytically.

Whittle index policy is designed for **Multi-source Scheduling.**

Numerical Result



8 times performance gain.

References

[1] MKC Shisher and Y Sun, "How Does Data Freshness Affect Real-time Supervised Learning", ***ACM Mobihoc***, 2022.

[2] Y Sun and B Cyr, "Sampling for Data Freshness Optimization: Non-linear Age Functions", ***JCN,*** 2019.

[3] MKC Shisher, H Qin, L Yang, F Yan, and Y Sun, "The Age of Correlated Features in Supervised Learning Based Forecasting", ***IEEE Infocom AoI Workshop.,*** 2021.

[4] J Gittins, K Glazebrook, and R Weber. " Multi-armed bandit allocation indices". ***John Wiley & Sons***, 2011.