Ph.D. Dissertation Defense

Timely Inference over Networks

Md Kamran Chowdhury Shisher

Department of Electrical and Computer Engineering

Advisor: Prof. Yin Sun

Committee: Prof. Roy Yates
Prof. John Hung
Prof. Shiwen Mao
Prof. Thaddeus Roppel

Reader: Prof. Tao Shu

AUBURN UNIVERSITY

April 3rd, 2024

Personal Info

- Education Background
  - 2013-2017 B.Sc. in Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
  - 2022 M.S. in Electrical Engineering, Auburn University
  - 2018-present Ph.D. in Electrical Engineering, Auburn University

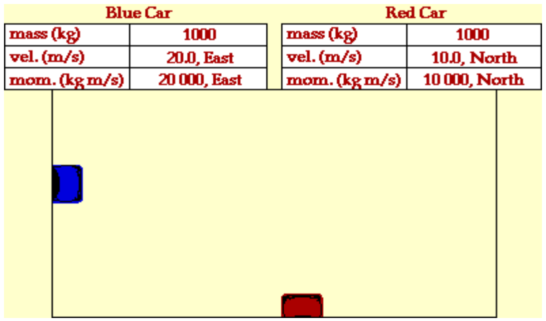
Outline

- Introduction and Motivation
  - Remote inference for networked intelligence
  - Metric for data freshness: Age of Information

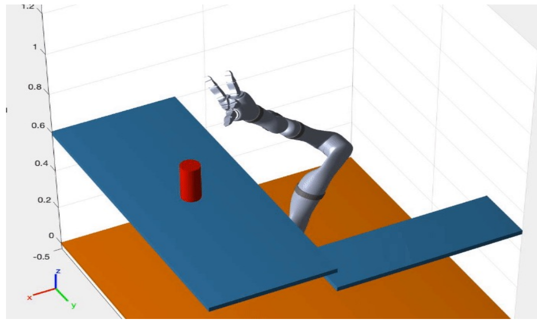
- Two topics that we will discuss:
  - How does data freshness affect remote inference or estimation?
    - A counter-intuitive phenomenon: Fresher data is not always better
    - INFOCOM Workshop 2021, MobiHoc 2022, INFOCOM Workshop 2024
  - How to improve inference performance?
    - We design feature length selection strategies and scheduling policies
    - MobiHoc 2022, JSAIT 2023, ToN 2023 (Under Review), ISIT 2024 (Under Review)
- Summary
- Other Works
  - MILCOM 2023, AgriEngineering 2023

Networked Intelligent Systems

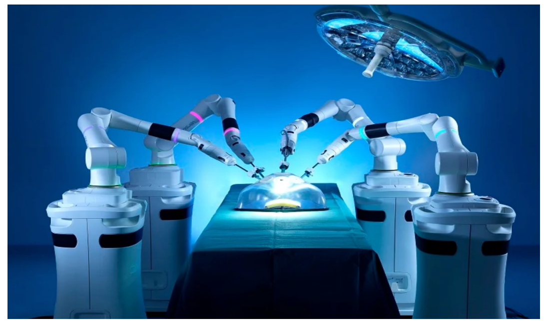
Vehicular Network



Factory Automation

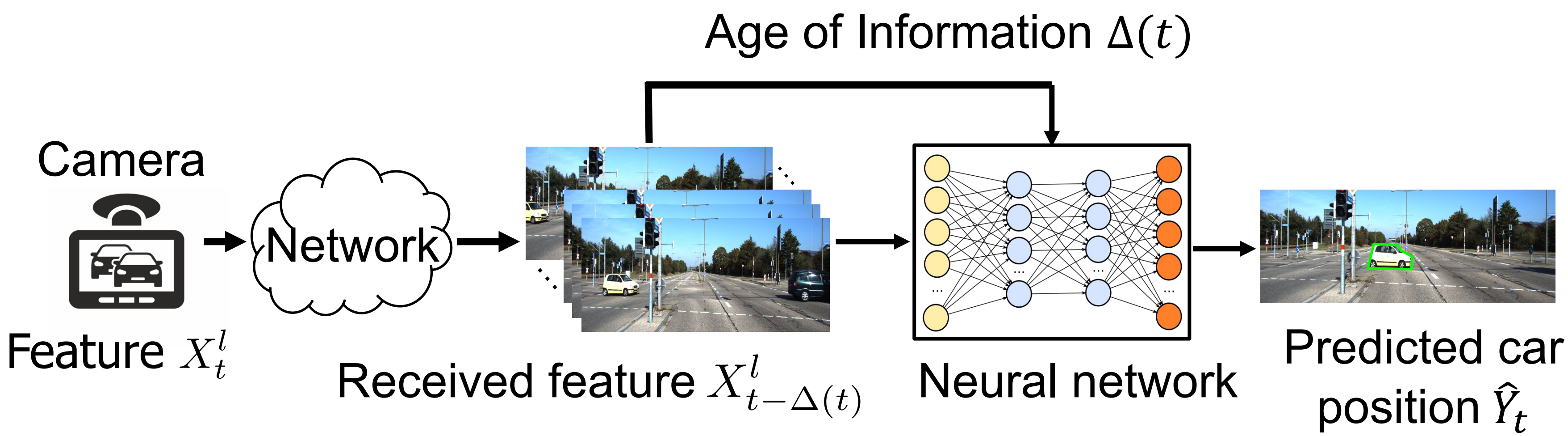


Remote Surgery



- Other applications of networked intelligence:
  - UAV/robot navigation, augmented/virtual reality, real-time surveillance, video analytics, etc.
  - Inference/estimation is crucial
- Autonomous vehicle predicts the trajectories of nearby vehicles
- In remote surgery, doctor needs to know movements of surgical robot

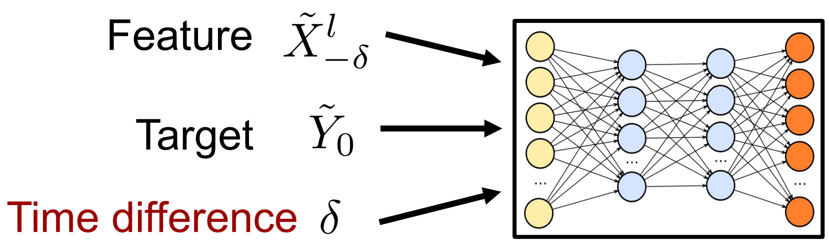
Remote Inference System



- Due to communication delay, delivered features may not be fresh
- At time t, the freshest received feature $X_{t-\Delta(t)}^l = (V_{t-\Delta(t)}, V_{t-\Delta(t)-1}, \ldots, V_{t-\Delta(t)-l+1})$ was generated at time $t - \Delta(t)$
- Time difference $\Delta(t)$ is the staleness of feature $X_{t-\Delta(t)}^l$ for inferring target Y_t at time t
- $\Delta(t)$ → Age of Information (AoI)

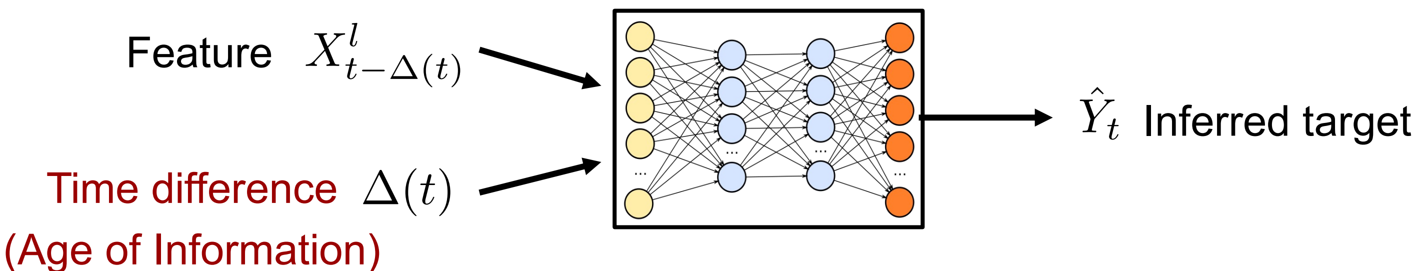
Offline Training and Online Inference

- Offline training
  - Neural network is trained offline using pre-prepared training dataset $(\tilde{Y}_0, \tilde{X}^l_{-\delta}, \delta)$
  - Time difference δ may vary across training data samples
  - We used Empirical Risk Minimization-based Supervised Learning algorithms



Neural network

- Online inference
  - We fed Feature and AoI into the neural network to infer target



Trained neural network

Loss Function

- Let ϕ_l^* be a trained neural network
- Loss function: $L(y, a)$ if $Y_t = y$, and output is $a = \phi_l^*\left(\Delta(t), X_{t-\Delta(t)}^l\right) \in \mathcal{A}$
- Loss function $L(\cdot,\cdot)$ is chosen based on goal of application
- Softmax regression, Log loss $L\left(y, Q_y\right) = -\log Q_Y(y)$
- Minimum mean squared estimation, Quadratic loss $L(y, \hat{y}) = ||y - \hat{y}||_2^2$

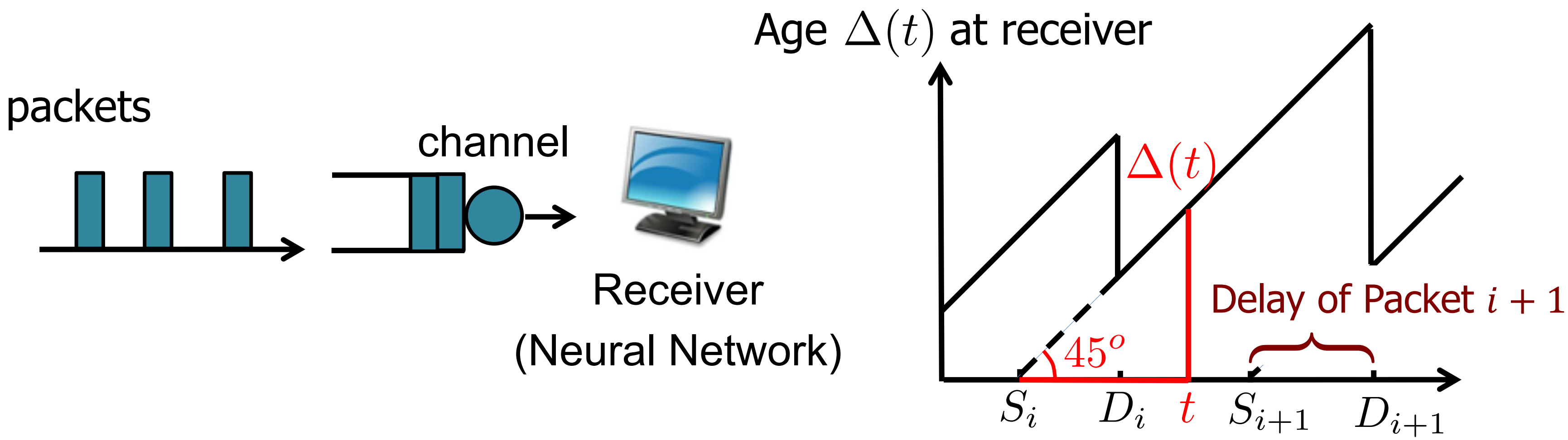
Inference Error

- The inference error is the expected loss over the joint distribution of target Y_t and feature $X^l_{t-\delta}$.
- Given AoI $\Delta(t) = \delta$ and feature sequence length l, inference error can be expressed as a function of AoI δ and feature length l:

$$\mathrm{err}_{\mathrm{inference}}(\delta, l) = \mathbb{E}_{Y, X^l \sim P_{Y_t, X^l_{t-\delta}}}\left[L(Y, \phi^*_l(\delta, X^l))\right]$$

  - ϕ^*_l is a trained neural network
  - $P_{Y_t, X^l_{t-\delta}}$ is the joint distribution of target Y_t and feature $X^l_{t-\delta}$

Age of Information: Definition



Definition: At time t, the **Age of Information (AoI)** $\Delta(t)$ is time difference between the current time t and the generation time of the freshest received packet

- If packet i is generated at S_i and delivered at D_i

$$\Delta(t) = \underbrace{t - \max\{S_i : D_i \le t\}}_{\text{Time difference between data generation and usage}}$$

Related Work

- Analysis and Optimization:
  - Average and Peak AoI [Kaul et al, ISIT 2012], [Yates, ISIT 2015], [Sun, TIT 2017].
  - Monotonic functions AoI [Sun et al, SPAWC 2018, JCN 2019], [Ornee and Sun, WiOpt 2019, ToN 2021], [Sun, TIT 2017], [Klugel et al, INFOCOM Workshop 2019]
  - Scheduling and Sampling Policies to minimize Average AoI, Peak AoI, and increasing functions of AoI [Sun et al, SPAWC 2018, JCN 2019], [Kadota et al, ToN 2018], [Kadota et al, INFOCOM 2018], [Klugel et al, INFOCOM Workshop 2019], [Tripathi et al, Allerton 2019]
- Lack of understanding regarding how to evaluate the importance of fresh information in real-time systems

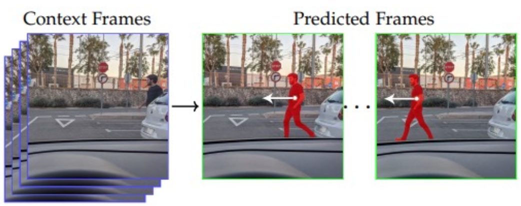
Research Questions

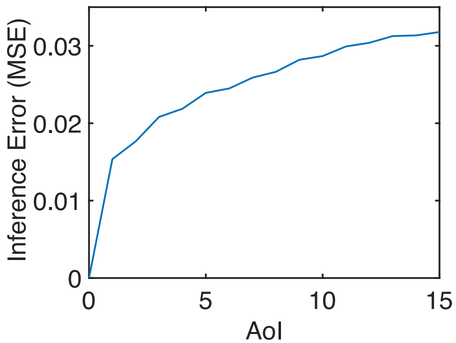
- Q1. Is Fresher data always better?
  - We propose a new information-theoretic tool to evaluate the importance of fresh information on remote inference system
- Q2. How to improve inference performance?
  - We designed scheduling policies based on the insights from information theoretic analysis

Q1: Is Fresher data better?

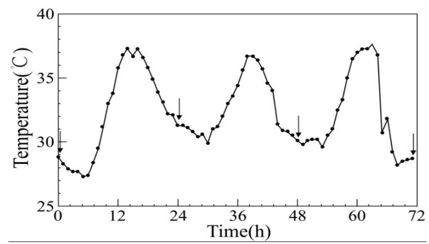
- Earlier studies assumed system performance degrades monotonically as feature becomes stale
- Not always true!

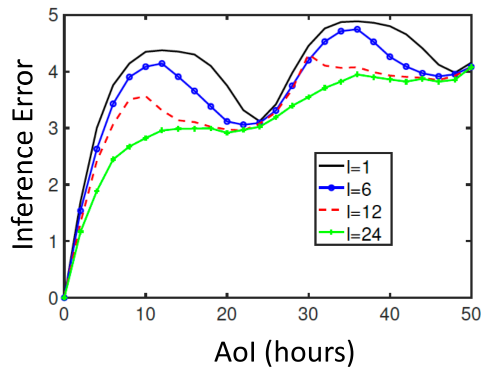
Video Prediction



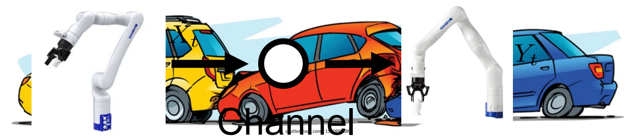


Temperature Prediction



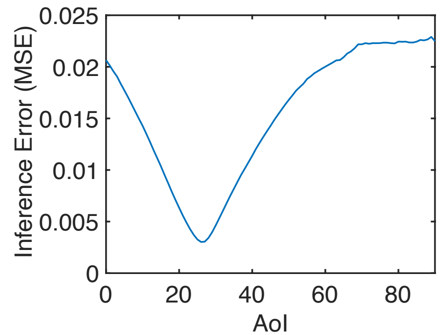


Reaction Prediction



Leader robot

Follower robot



Fresh data is Better

Fresher data is not always Better

Inference Error

- Because neural network function is complicated, it is difficult to interpret the experimental observations!

$$\mathrm{err}_{\mathrm{inference}}(\delta, l) = \mathbb{E}_{Y,X^l \sim P_{Y_t, X_{t-\delta}^l}}\left[L\left(Y, \phi_l^*\left(\delta, X^l\right)\right)\right]$$

Information-theoretic Metric for Inference Error

- L-Entropy [F. Farnia and D. Tse, NeurIPS, 2016]: The L-entropy of a random variable Y is the minimum expected loss for predicting Y:

$$H_L(Y) = \min_{a \in \mathcal{A}} \mathbb{E}_{Y \sim P_Y}[L(Y, a)]$$

- L-Conditional Entropy: Given X, the minimum expected loss for predicting Y is L-Conditional Entropy $H_L(Y|X)$:

$$H_L(Y|X) = \sum_{x \in \mathcal{X}} P_X(x)\, H_L(Y|X = x)$$

$$H_L(Y|X = x) = \min_{a \in \mathcal{A}} \mathbb{E}_{Y \sim P_{Y|X=x}}[L(Y, a)]$$

- If L is log loss, then $H_L(Y|X)$ is Shannon Conditional Entropy
- If L is quadratic loss, then $H_L(Y|X)$ is minimum mean square error

Farzan Farnia and David Tse, "A minimax approach to supervised learning," NeurIPS, 2016.

Information-theoretic Metric for Inference Error

- Inference error is lower bounded by L-Conditional entropy

$$\begin{aligned}\mathrm{err}_{\mathrm{inference}}(\delta, l) &= \mathbb{E}_{Y,X^l \sim P_{Y_t, X^l_{t-\delta}}}\left[L\left(Y, \phi_l^*\left(\delta, X^l\right)\right)\right] \\ &\geq \min_{\phi_l \in \Phi_l} \mathbb{E}_{Y,X^l \sim P_{Y_t, X^l_{t-\delta}}}\left[L\left(Y, \phi_l\left(\delta, X^l\right)\right)\right] \\ &= H_L\left(Y_t \middle| X^l_{t-\Delta(t)}\right),\end{aligned}$$

where Φ_l is the set of all functions mapping from input space to output space.

$$\mathrm{err}_{\mathrm{inference}}(\delta, l) = H_L(Y_t | X^l_{t-\Delta(t)}) + \mathrm{O}(\beta),$$

where β is the KL-divergence between the distributions of the training dataset and inference dataset

- It is convenient to analyze $H_L(Y_t | X^l_{t-\Delta(t)})$ than $\mathrm{err}_{\mathrm{inference}}(\delta, l)$

L-Conditional Entropy vs. AoI

- Data Processing Inequality for Markov chain: [A. Dawid'98]

Lemma: If $Y_t \leftrightarrow X_{t-\mu}^l \leftrightarrow X_{t-\mu-\nu}^l$ is a Markov chain for all $\mu, \nu \geq 0$, then $H_L(Y_t|X_{t-\Delta(t)}^l)$ is an increasing function of AoI δ for a fixed l

- However, practical data in **Experiments** can be **non-Markovian**

A. P. Dawid, "Coherent measures of discrepancy, uncertainty and dependence, with applications to Bayesian predictive experimental design," Technical Report 139, 1998.

ϵ-Markov Chain

- Extend **data processing inequality** to **non-Markov chain**. How?

ϵ-Markov Chain

- Extend **data processing inequality** to **non-Markov chain**
  - Local information geometry method

    [Huang, Makur, Wornell, Zheng, 2019]

- **Definition:** [Shisher et al'21, Shisher & Sun'22, Shisher et al'24] Given $\epsilon \geq 0$, three random variables Y, X, and Z are said to be an **ϵ-Markov chain**, denoted as $Y \overset{\epsilon}{\leftrightarrow} X \overset{\epsilon}{\leftrightarrow} Z$ if
$$I_{Shannon}(Y; Z|X) = D_{KL}(P_{Y,X,Z} \,||P_{Y|X}P_{Z|X}P_X) \leq \epsilon^2.$$

- **ϵ-Markov chain** measures divergence of a data sequence from a Markov chain.
- If $\epsilon = 0$, then ϵ-Markov chain $Y \overset{\epsilon}{\leftrightarrow} X \overset{\epsilon}{\leftrightarrow} Z$ is also a Markov chain

Huang, Makur, Wornell, Zheng, "On Universal Features for High-Dimensional Learning and Inference," ArXiv, 2019.
Shisher et al, "The Age of Correlated Features in Supervised Learning based Forecasting," AoI Workshop 2021.
Shisher and Sun, "How Does Data Freshness Affect Real-time Supervised Learning?" ACM MobiHoc 2022.
Shisher et al, "Timely communications for remote inference," 2024.

A1: Fresher is Not always Better

- **ϵ-Data Processing Inequality** [Shisher et al'21, Shisher & Sun'22]

Theorem: If $Y_t \overset{\epsilon}{\leftrightarrow} X_{t-\mu}^{l} \overset{\epsilon}{\leftrightarrow} X_{t-\mu-\nu}^{l}$ is an ϵ-Markov chain for all $\mu, \nu \geq 0$, then

$$H_L\left(Y_t \middle| X_{t-\delta}^{l}\right) = g_l(\delta) + O(\epsilon)$$

where $g_l(\delta) = \sum_{k=0}^{\delta-1} I_L(Y_t; X_{t-k} | X_{t-k-1})$ is an increasing function of AoI δ

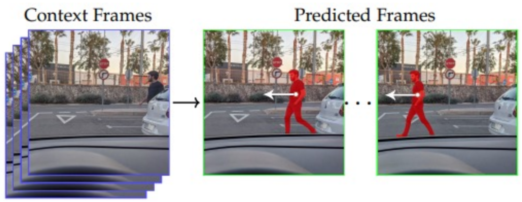
- If $\epsilon = 0$, $H_L(Y_t | X_{t-\delta}^{l})$ is an increasing function of AoI δ.
- If ϵ is far from 0, then $H_L(Y_t | X_{t-\delta}^{l})$ may not be monotonic with AoI

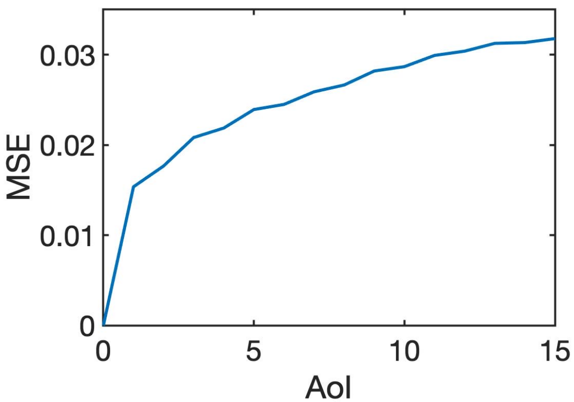
- For Markovian data, fresher data ➔ smaller error
- For non-Markovian data, fresher data is NOT always better

Shisher et al, "The Age of Correlated Features in Supervised Learning based Forecasting," AoI Workshop 2021.
Shisher and Sun, "How Does Data Freshness Affect Real-time Supervised Learning?" ACM MobiHoc 2022.

Lesson Learned: Data Freshness for Remote Inference

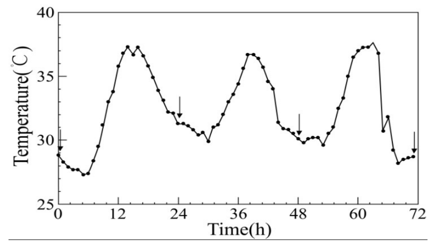
Video Prediction

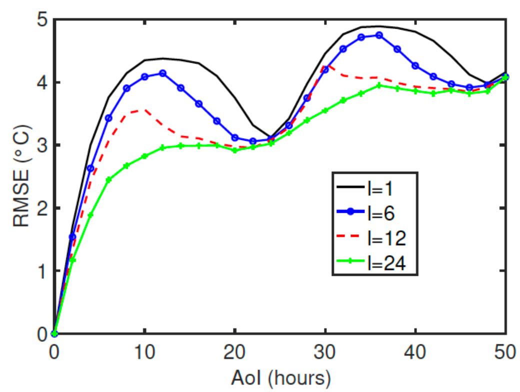




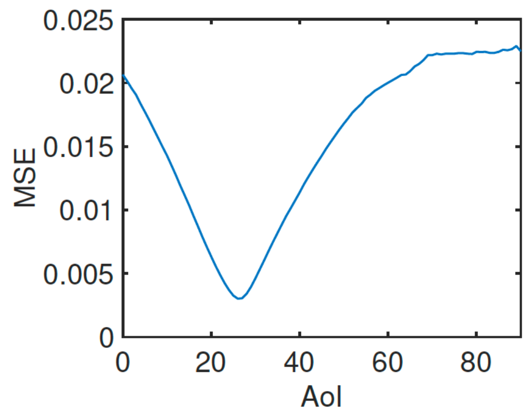
Markov data sequence

Temperature Prediction





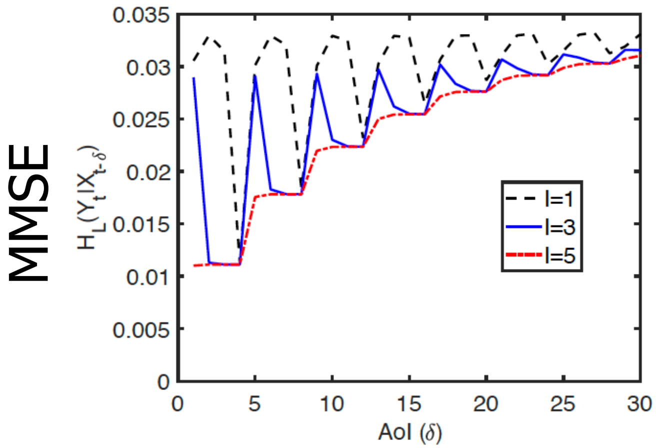
Reaction Prediction



Non-Markov data sequences

- In remote inference, inference error is a function of AoI.
  - For Markovian data, fresher data ➔ smaller error
  - For non-Markovian data, fresher data is NOT always better
  - Tools: Information theory and local information geometry

Analytical Model: When Fresh Data is Better?



- AR(4) signal: $V_t = 0.1V_{t-1} + 0.8\,V_{t-4} + W_t$
  - use $X_{t-\delta}^{l} = (V_{t-\delta}, V_{t-\delta-1}, \ldots, V_{t-\delta-l+1})$ with length l to estimate $Y_t = V_t + N_t$
  - $\epsilon(l)$: $\epsilon(1) = 1.6, \epsilon(3) = 1.3,$ and $\epsilon(5) = 0$

As feature length l increases

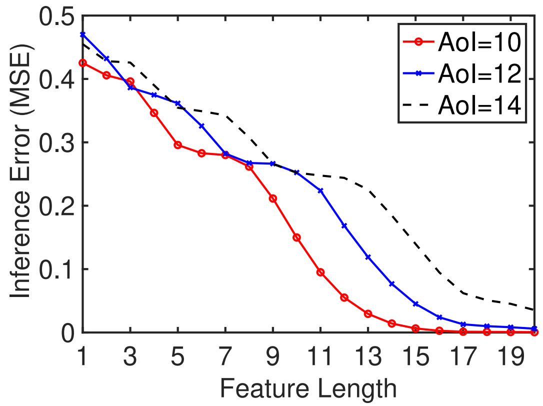
- Parameter ϵ in ϵ-Markov chain $V_t \overset{\epsilon}{\leftrightarrow} X_{t-\mu}^{l} \overset{\epsilon}{\leftrightarrow} X_{t-\mu-\nu}^{l}$ decreases to 0
- MMSE tends to an increasing function of AoI δ

Q2. How to improve Inference performance?

- Use inference error as the performance metric to design timely communication protocols for remote inference system
- We design **optimal** feature selection strategies and scheduling policies

Observations

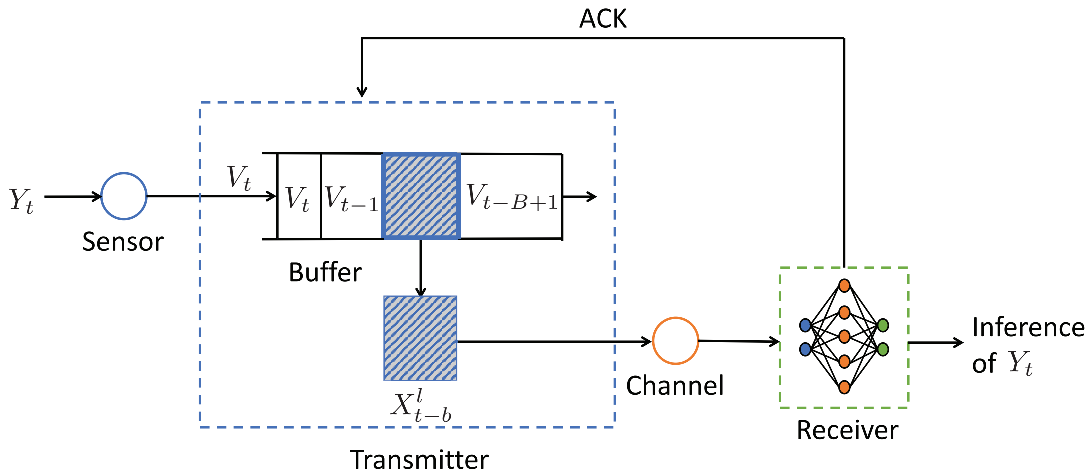
- For non-Markovian data, fresher data is NOT always better
- Longer feature reduces inference error. See:



Channel state information prediction (Jakes Model)

- But Longer feature requires more communication resources
- A longer feature can take more time slots to reach destination and ends up being stale which may result in worse inference

New “Selection-from-Buffer” Model



- Buffer stores B most recent sensor measurements $(V_{t,} V_{t-1}, \dots, V_{t-B+1})$
- When to send? Which packet to send?
- **Scheduler** sends $X^l_{t-b} = (V_{t-b,} V_{t-b+1}, \dots, V_{t-b-l+1})$
- Feature length l determines Longer or Shorter?
- Starting position b of feature in buffer is feature position-Fresh or Stale?
- More general than “Generate-at-Will” model [Yates’2015, Sun et al.’2017]
  - When B=1, “Selection-from-Buffer” model reduces to “Generate-at-Will” model

Roy Yates, “Lazy is timely: Status updates by an energy harvesting source,” ISIT 2015.
Sun et al, “Update or Wait: How to Keep Your Data Fresh ,” TIT 2017.

Scheduling Problem with Fixed Feature Length

- Scheduling problem:

$$\bar{p}_{l,opt} = \inf_{\pi \in \Pi} \limsup_{T \to \infty} \mathbb{E}_{\pi}\left[\frac{1}{T}\sum_{t=0}^{T-1} p(\Delta(t), l)\right]$$

$p(\Delta(t), l) = \mathrm{err}_{\mathrm{inference}}\,(\Delta(t), l)$ is the actual inference error, which can be either monotonic or non-monotonic with AoI.

- Scheduling policy with fixed feature length l: $\pi = (f, g)$
  - $g = (S_1, S_2, \ldots)$ decides sending times
  - $f = (b_1, b_2, \ldots)$ decides the feature position (fresh or stale)
- **Difficulty:**
  - Non-monotonic AoI function.
  - Need to determine both feature position and feature sending times.
  - Earlier studies assumed monotonic AoI functions and determined only feature sending times [Sun et al'21, Klugel et al'19]

Sun et al, "Sampling for Data Freshness Optimization: Non-linear Age Functions," JCN 2021.
Klugel et al, "AoI-Penalty Minimization for Networked Control Systems with Packet Loss," AoI Workshop 2019.

Sub-Scheduling Problem

- Fix $f_b = (b, b, \dots)$, i.e., always send the b-th oldest packet in the buffer, then optimize the feature sending times $g = (S_1, S_2, \dots)$

$$\bar{p}_{l,b} = \inf_{(S_1, S_2, \dots)} \limsup_{T\to\infty} \mathbb{E}_{(f_b, g)}\left[\frac{1}{T}\sum_{t=0}^{T-1} p(\Delta(t), l)\right]$$

Theorem: It is optimal to send the next feature when

(i) the channel is idle and

(ii) an index exceeds threshold β: $\gamma_l\big(\Delta(t)\big) \geq \beta$,

where β is the unique root of a fixed point equation [Bisection Search]

$$\mathbb{E}\left[\sum_{t=D_i(\beta)}^{D_{i+1}(\beta)-1} p(\Delta(t), l)\right] - \beta\, \mathbb{E}[D_{i+1}(\beta) - D_i(\beta)] = 0$$

and D_i is the delivery time of i-th feature. Moreover $\beta = \bar{p}_{l,b}$.

Tool: Restart-in-Random-State Formulation

Restart-in-Random-State Problem

- A feature can take random time slots $T_i(l)$ for transmission
- A new feature delivers → AoI starts from a random state $b + T_i(l)$
- Restart-in-Random-State problem: When to restart from a random state?
- In literature, there exists Restart-in-State problem
  - Gittins index-based threshold policy [Katehakis and Veinott,1987]
- We obtain the Bellman eqn. of our Restart-in-Random-State problem:

$$h(\delta) = \inf_{\tau \in \{0,1,2,\dots\}} \mathbb{E}\left[\sum_{k=0}^{\tau+T_1(l)-1} \left(p(\delta+k,l) - \bar{p}_{l,b}\right)\right] + \mathbb{E}\left[h\left(b+T_1(l)\right)\right]$$

- Index function is $\gamma_l(\delta) = \inf_{\tau \in \{1,2,\dots\}} \frac{1}{\tau} \mathbb{E}\left[\sum_{k=0}^{\tau} p(\delta + T_1(l) + k)\right]$

First work to connect AoI Optimization with Restart-in-Random-State problem!

Katehakis and Veinott, "The Multi-Armed Bandit Problem: Decomposition and Computation," Math. of Operations Research, 1987

A2: Optimal Scheduling Policy with Fixed Feature Length

- Scheduling policy $\pi = (f, g)$, where $f = (b_1, b_2, \ldots)$ and $g = (S_1, S_2, \ldots)$

$$\bar{p}_{l,opt} = \inf_{\pi \in \Pi} \limsup_{T \to \infty} \mathbb{E}_\pi \left[\frac{1}{T} \sum_{t=0}^{T-1} p(\Delta(t), l) \right]$$

Theorem: The optimal scheduling policy

(1) always sends the b^*-th oldest packet in the buffer, where

$$b^* = \arg \min_{b \in \{0,1,\ldots,B-1\}} \bar{p}_{l,b}$$

(2) sends the next feature when (i) the channel is idle and (ii) the index $\gamma_l(\Delta(t))$ exceeds threshold μ, where

$$\gamma_l(\delta) = \inf_{\tau \in \{1,2,\ldots\}} \frac{1}{\tau} \mathbb{E}\left[\sum_{k=0}^{\tau} p(\delta + T_1(l) + k)\right]$$

Moreover,

$$\mu = \bar{p}_{l,opt} = \min_{b \in \{0,1,\ldots,B-1\}} \bar{p}_{l,b}$$

Shisher and Sun, "How Does Data Freshness Affect Real-time Supervised Learning?" ACM MobiHoc 2022.

Shisher et al, "Timely communications for remote inference," 2024.

A2: Optimal Scheduling Policy: with Time-varying Feature Length

- Scheduling policy: $\pi = ((S_1, b_1, l_1), (S_2, b_2, l_2), \ldots)$

$$\bar{p}_{opt} = \inf_{\pi \in \Pi} \limsup_{T \to \infty} \mathbb{E}_\pi \left[\frac{1}{T} \sum_{t=0}^{T-1} p(\Delta(t), d(t)) \right]$$

where $d(t)$ is feature length of the freshest feature at receiver

Theorem: The optimal feature length l_{i+1}^* is obtained by using the following Bellman equation:

$$h(\delta, d) = \min_{l \in \{1,2,\ldots,B\}} \mathbb{E}\left[\sum_{k=0}^{Z_l(\delta,d)+T_1(l)-1} \left(p(\delta + k, l_i^*) - \bar{p}_{opt}\right) \right] + \mathbb{E}\left[h\big(T_1(l) + b(l)\big)\right]$$

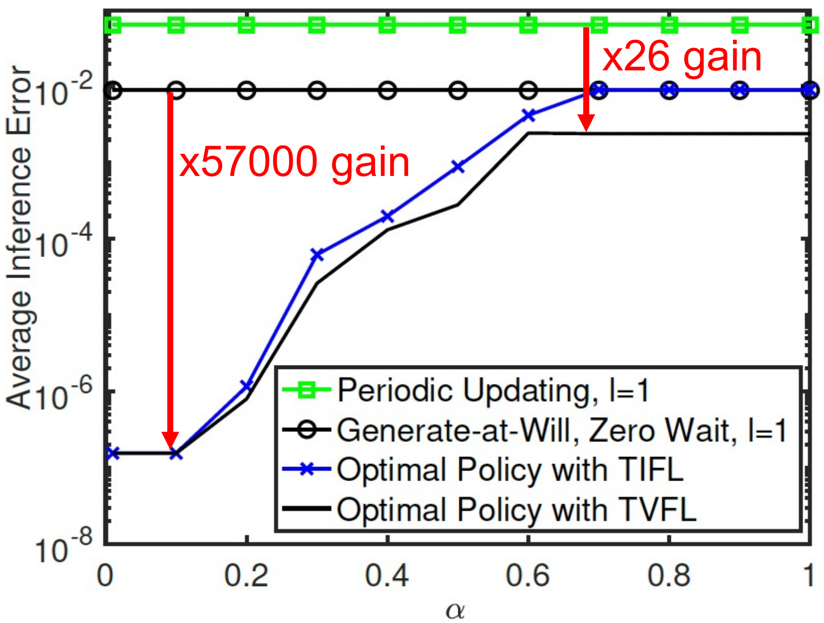
where $h(\cdot)$ is the relative value function, $Z_l(\delta, d)$ and $b(l)$ are the solution to the optimal scheduling policy with feature length l.

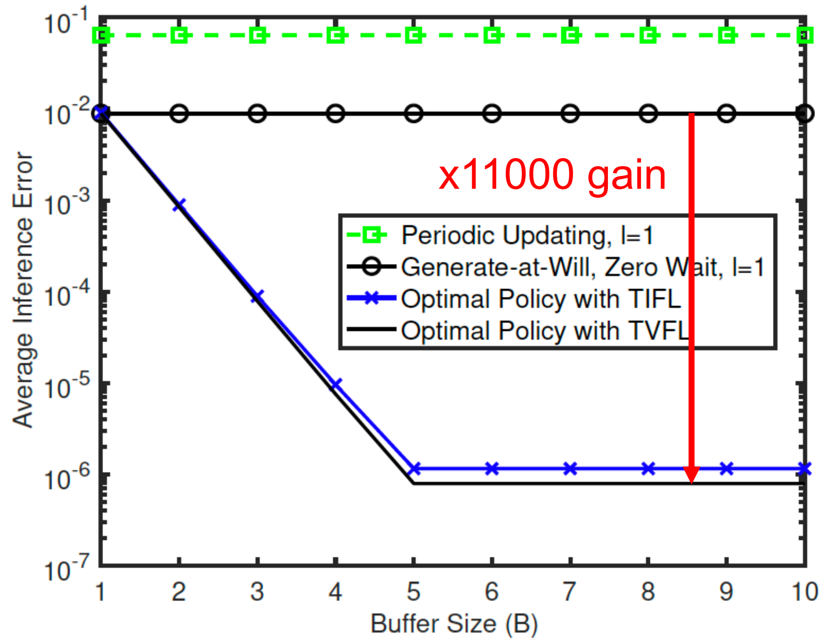
- Simplified policy iteration algorithm is provided in [Shisher et al'23]

Shisher, Bo Ji, I-Hong Hou, and Sun, "Learning and Communications Co-Design for Remote Inference Systems: Feature Length Selection and Transmission Scheduling," *IEEE JSAIT*, 2023.

.

Performance: Inference Error Reduction

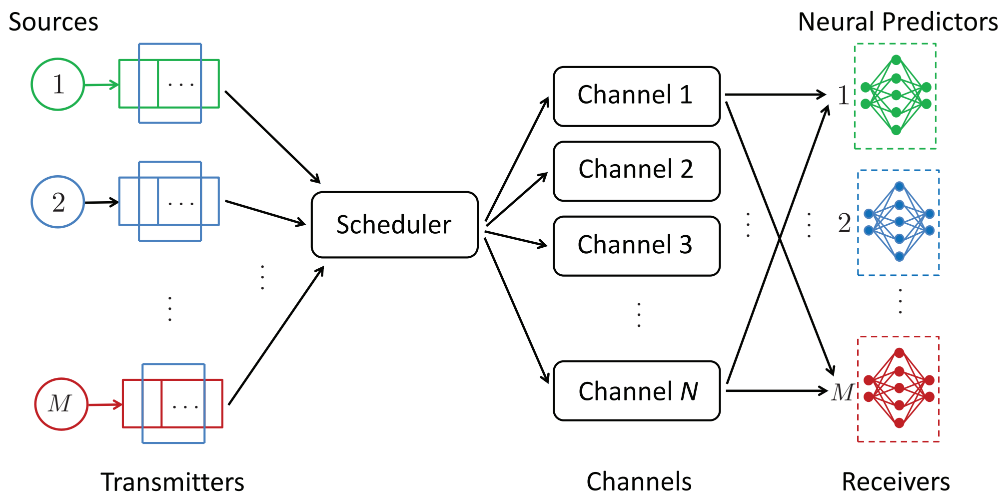




- Experiment: Channel state information prediction
- Transmission time $T(l) = \lceil \alpha l \rceil$, where l is feature length

Optimal scheduling policy with feature length selection reduces the inference error by up to 10000 times

Q3. Multi-Source, Multi-Channel Scheduling?



- Each source m stores B_m most recent features
- Limited number of communication channel
- At time slot t, a centralized scheduler decides:
  - which sources to select and which feature from the sources buffer to send
- The feature transmission times are random and non-preemptive
- A source can be served by at most one channel at a time

Scheduling Problem with Fixed Feature Length

- Scheduling Problem:

$$\text{minimize} \limsup_{T\to\infty} \sum_{m=1}^{M} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} p_m(\Delta_{\mathrm{m}}(t), l_m)\right]$$

$$\text{s.t.} \sum_{m=1}^{M} c_m(t) \le N,\ \ c_m(t) \in \{0,1\}, t = 0, 1, \ldots, m = 1, \ldots, M$$

where $c_m(t)$ is the channel occupation status of source m.

- Scheduler decides: Which sources to select? Which feature (fresh or stale) from selected source to send?
- Restless Multi-armed Bandit Problem (RMAB) with multiple actions
- PSPACE-hard
- Earlier studies [Kadota et al, ToN 2018], [V. Tripathi et al, Allerton 2021]
  - Monotonic AoI function and Binary action
  - Utilizes Whittle Index
- New policy integrating Whittle index with duality-based feature selection

Equivalent Problem

- Main Problem:

$$\text{minimize} \limsup_{T\to\infty} \frac{1}{T}\sum_{m=1}^{M} \mathbb{E}[\sum_{t=0}^{T-1} p_m(\Delta_\mathrm{m}(t), l_m)]$$

$$\text{s.t.} \sum_{m=1}^{M} c_m(t) \le N,\ \ c_m(t) \in \{0,1\}, t = 0, 1, \ldots, m = 1, \ldots, M$$

- Standard RMAB considers equality constraint
- To fit with standard RMAB problem, we add N dummy bandits
- Each Dummy Bandit:
  - Occupies one channel for one time-slot if scheduled
  - Incurs no cost

$$\text{minimize} \limsup_{T\to\infty} \frac{1}{T}\sum_{m=0}^{M} \mathbb{E}[\sum_{t=0}^{T-1} p_m(\Delta_\mathrm{m}(t), l_m)]$$

$$\text{s.t.} \sum_{m=0}^{M} c_m(t) = N,\ \ c_m(t) \in \{0,1\}, t = 0, 1, \ldots, m = 1, \ldots, M$$

$$c_0(t) \in \{0, 1, \ldots, N\}, t = 0, 1, \ldots.$$

Decoupled Problem

- Constraint Relaxation:
- Per time slot constraint to time-average constraint

$$\limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}\sum_{m=0}^{M} c_m(t)\right] = N, c_m(t) \in \{0,1\}, c_0(t) \in \{0,1,\dots,N\}$$

- Lagrangian Dual Decomposition: $(M+1)$-decoupled problems
- M per-source problems, where m-th source problem is

$$\text{minimize} \limsup_{T\to\infty} \mathbb{E}_\pi\left[\frac{1}{T}\sum_{t=0}^{T-1} p_m(\Delta_\mathrm{m}(t), l_m) + \lambda\, c_m(t)\right]$$

- Dual variable λ→ one time slot channel occupancy cost
- One problem for dummy bandits:

$$\text{minimize} \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} \lambda\, c_0(t)\right]$$

Whittle Index for Source Selection

- Difficulty:
  - Evaluating Whittle index requires closed-form solution to the per-source decoupled problem
  - An indexability condition needs to be satisfied
  - Indexability requires a nice solution structure to the per-source decoupled problem

Indexability

- Each m-th decoupled problem is a MDP:
- δ: AoI of source m
- β: amount of time used to send the current feature from source m
  - $\beta = 0 \Longrightarrow$ no feature from source m is currently in service
  - $\beta > 0 \Longrightarrow$ a feature from source m is currently in service

Passive Set: $\Psi_m(\lambda)$ is the set of states $(\delta, \beta) \in \mathbb{Z}^+ \times [0, \infty)$ such that if $\Delta(t) = \delta$ and $\beta(t) = \beta$, then it is optimal to not select source m.

Indexability: The m-th decoupled problem is indexable if $\Psi_m(\lambda)$ increases with the increase of the cost λ from $-\infty$ to ∞, i.e., $\lambda_1 \leq \lambda_2$ implies $\Psi_m(\lambda_1) \subseteq \Psi_m(\lambda_2)$.

- Because solution to the per-source problem follows an index-based threshold policy for a channel occupancy cost λ, we were able to simplify

$$\Psi_m(\lambda) = \{(\delta, \beta): \beta > 0 \text{ or } \gamma_{l,m}(\delta) < \mu(\lambda)\}$$

Theorem: All decoupled problems are indexable

Whittle Index

- Whittle Index: $W_m(\delta, \beta) = \inf\{\lambda : (\delta, \beta) \in \Psi_m(\lambda)\}$
- Whittle index is the level of channel occupancy cost λ such that selecting and not selecting the source m is equality profitable.

Theorem: Whittle index $W_m(\delta, \beta)$ for each source is

(a) If $\beta > 0$: $W_m(\delta, \beta) = -\infty,\ m = 1, \dots, M$

(b) if $\beta = 0$: We obtained closed-form expression from

$$W_m(\delta, 0) = \inf\{\lambda : \gamma_{l,m}(\delta) < \mu(\lambda)\},\ m = 1, \dots, M$$

(c) For dummy bandits, Whittle index $W_0(\delta, \beta) = 0$

- First work to obtain Whittle index for
  - Non-monotonic AoI function
  - Random delay channel
  - Multi-action problem

Shisher and Sun, "How Does Data Freshness Affect Real-time Supervised Learning?" ACM MobiHoc 2022.

Shisher et al, "Timely communications for remote inference," 2024.

A3. Proposed Policy with Fixed Feature Length

- For all time-slot t:
  - Select sources with maximum Whittle indices
  - Sources with negative Whittle index is never selected
  - If source m is selected, send b_m^*-th oldest feature
    - b_m^* is the optimal feature position of the decoupled problem

      minimize $\limsup_{T\to\infty} \mathbb{E}_\pi \left[\frac{1}{T}\sum_{t=0}^{T-1} p_m(\Delta(t), l_m) + \lambda^* \, c_m(t)\right]$
    - λ^* is the optimal Lagrange Multiplier of a dual problem
    - Stochastic Sub-gradient Ascent Algorithm

Theorem: Our proposed policy is asymptotically optimal as the number of sources and channels increases to infinity, maintaining a constant ratio

Tool: Linear Programming-based priority conditions [Gast et al, 2023]

Nicolas Gast, Bruno Gaujal, Chen Yan, "Linear Program-Based Policies for Restless Bandits: Necessary and Sufficient Conditions for (Exponentially Fast) Asymptotic Optimality," Mathematics of Operations Research, 2023.

.

A3. Proposed Policy with Time-varying Feature Length

- Difficulty:
  - Indexability is difficult to establish
  - We can not have Whittle index
  - Proposed a new policy that does not require indexability

Maximum Gain First (MGF) Policy [Shisher et al, JSAIT, 2023]

- For all time-slot t:
  - Select sources and features with maximum positive Gain indices

    Gain index=Action value for not selecting − Action value for selecting

Shisher, Bo Ji, I-Hong Hou, and Sun, "Learning and Communications Co-Design for Remote Inference Systems: Feature Length Selection and Transmission Scheduling," *IEEE JSAIT*, 2023.

.

Summary

- **Timely Inference over Networks**
  - **Idea:** Use inference error as the performance metric to design timely communication protocols
- **Inference Error = f(AoI, feature length)**
  - Markovian data ➔ fresher data is better
  - non-Markovian data ➔ fresher data is NOT necessarily better
  - Longer Feature ➔ Better inference
  - Longer Feature ➔ More communication resources
- **Communications for Remote Inference**
  - New “Selection-from-Buffer” model
  - Optimal Single-source Single-channel scheduling policy
  - New Asymptotically optimal multi-source multi-channel scheduling policy
  - Proposed solution reduces inference error by 10000 times

Other Works and Services

- Status updating with two-way delay and non-*i.i.d*. channel
- Status updating for safety monitoring systems
- AI for Agriculture
  - USDA Funded project "AI-based Food Demand Forecasting for Alabama Food Pantry"
  - Transfer Learning-based disease detection in Banana crops.
- Teaching Assistant
  - Machine Learning
  - Reinforcement Learning
- Maintainer of an online paper repository on Age of Information: http://webhome.auburn.edu/~yzs0078/AoI.html

Publications

1. **M.K.C. Shisher**, Y. Sun, and I.-H. Hou, “Timely Communications for Remote Inference," under review in *IEEE/ACM ToN, 2023*.
2. **M.K.C. Shisher** and Y. Sun, “On the Monotonicity of Information Aging," *IEEE INFOCOM AoI workshop, 2024*.
3. C. Ari, **M.K.C. Shisher**, E. Uysal, and Y. Sun, “Goal-Oriented Communications for Remote Inference with Two-Way Delay," under review in IEEE ISIT, *2023*.
4. **M.K.C. Shisher**, B. Ji, I.-H. Hou, and Y. Sun, “Learning and Communications Co-Design for Remote Inference Systems: Feature Length Selection and Transmission Scheduling,” *IEEE JSAIT, 2023*.
5. **M.K.C. Shisher** and Y. Sun,“How Does Data Freshness Affect Real-time Supervised Learning?” *ACM MobiHoc, 2022* ***(Acceptance Rate: 19.8%)***
6. **M.K.C. 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.
7. K. Yan, **M.K.C. Shisher**, and Y. Sun, “A Transfer Learning-Based Deep Convolutional Neural Network for Detection of Fusarium Wilt in Banana Crops,” *AgriEngineering*, 2023.
8. T. Z. Ornee, **M.K.C. Shisher**, C. Kam, and Y. Sun, “Context-aware Status Updating: Wireless Scheduling for Maximizing Situational Awareness in Safety-critical Systems,” *IEEE MILCOM*, 2023.

Future Works

- Wireless networks and information freshness
- Semantic Communications: How to design a semantic encoder and predictor/controller?
- RMAB with multiple actions
- Online scheduling problem for remote inference/control
- Reinforcement Learning-based Scheduler
- Implementation to UAV and Robots


Thank You!