

RANDOMIZED EDGE-ASSISTED ON-SENSOR INFORMATION SELECTION FOR BANDWIDTH-CONSTRAINED SYSTEMS

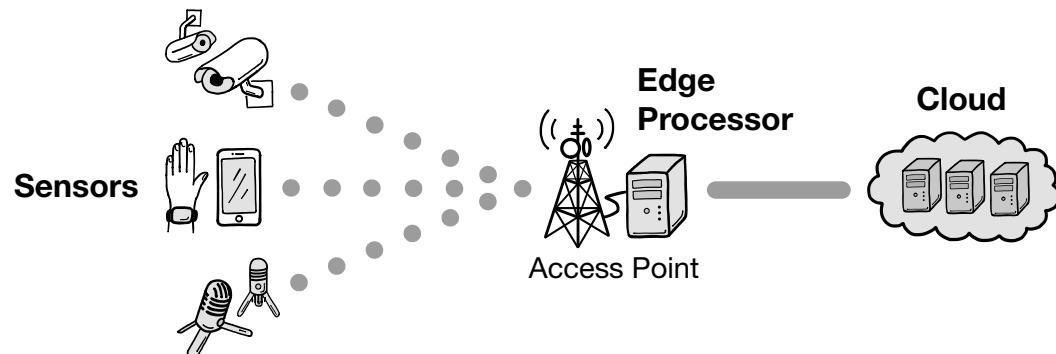
Igor Burago

Marco Levorato

Department of Computer Science
University of California, Irvine



INTELLIGENT IoT IN PRINCIPLE



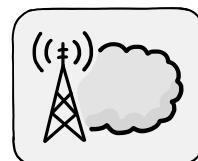
INFORMATION
Data to drive decisions



Sensor

- Observes the environment
- Selectively informs the edge
- Adjusts its own operation

INTELLIGENCE
Algorithms to realize autonomy



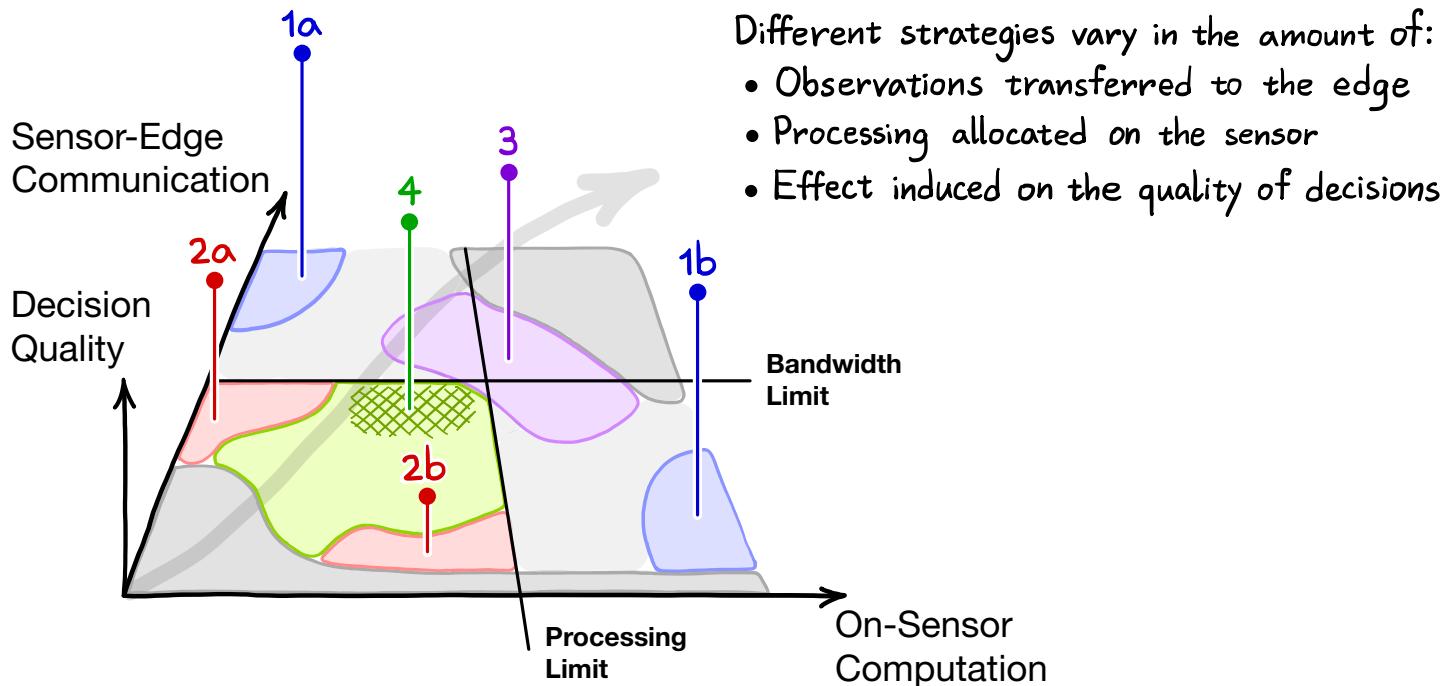
Edge/Cloud

- Accumulates relevant observations
- Exercises high-level control
- Assists on-sensor decision-making

Goal: Realize the alliance between
information completeness and best intelligence

INTELLIGENT IoT IN PRINCIPLE

Computation-communication tradeoffs



1a Offload all decisions to edge

1b Offload all decisions to sensor

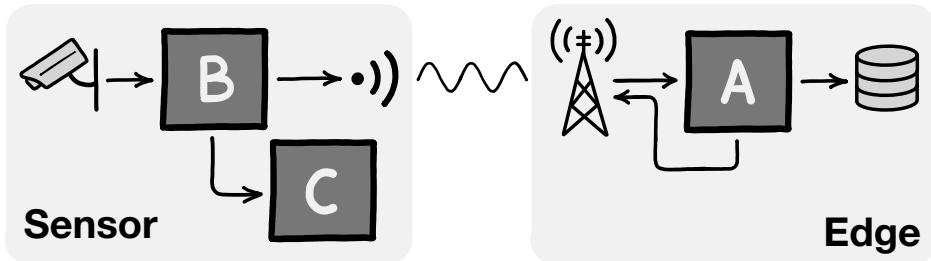
3 Transmit all in compressed form

2a Sample observations to transmit

2b Transmit obvious positives

4 Optimal adaptive compromise

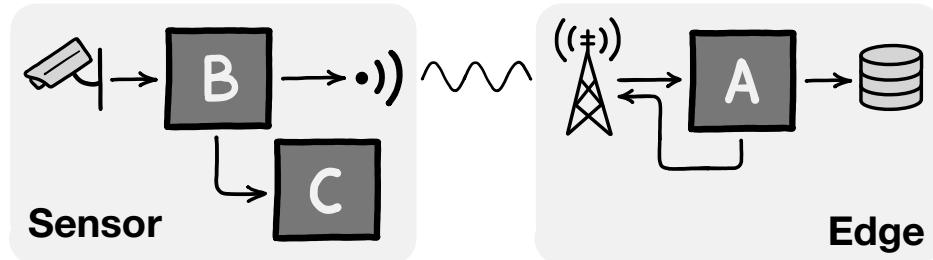
SENSOR-EDGE COOPERATION MODEL



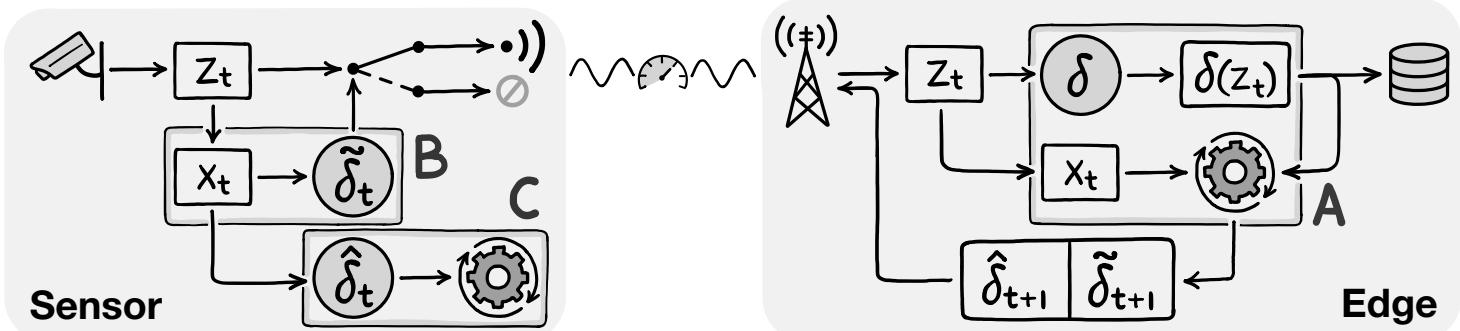
Model assumptions

- Observations fall into two classes:
 - of interest to the system (**positives**)
 - and not (**negatives**)
- Positives are to be collected at edge for further processing in the system
- Edge can run the optimal classifier δ ; sensor can not
- **B** Sensor runs a low-complexity classifier supplied by edge
- **A** Edge, via supervision of its classifier, trains the sensor's
- **C** Sensor uses its classifier to spare bandwidth by dropping some negatives, and to make local decisions

SENSOR-EDGE COOPERATION MODEL

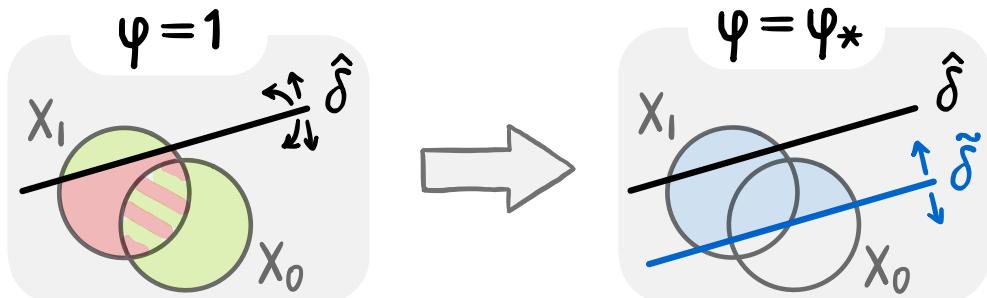
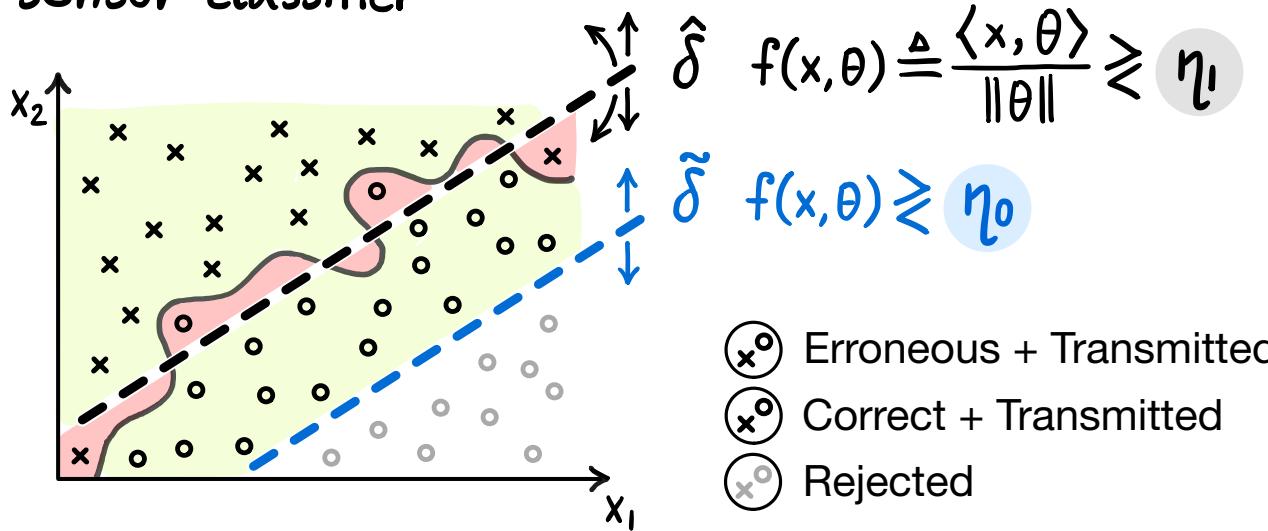


-
- Factors at play**
- CPU, battery, etc.
 - Computation constraints
 - Complexity of $\tilde{\delta}$ and $\hat{\delta}$
 - Shared wireless channel
 - Bandwidth constraint
 - Transmission probability $\varphi \approx \varphi_*$
 - Need to react to environment
 - Cost of error in feature space
 - Expected risk U to minimize
-



SENSOR-EDGE COOPERATION MODEL

On-sensor classifier



SENSOR-EDGE COOPERATION MODEL

Problem

$$U(\tau) \rightarrow \min, \text{ s.t. } W(\tau) = 0 \quad \Rightarrow \quad \frac{1}{\zeta} U(\tau) + W(\tau) \rightarrow \min_{\tau}$$

$$U(\tau) \triangleq \mathbb{E} \left[\hat{I}_{\eta_0}(u_\theta) I_e(z, \tau) |r_o(u_\theta - \eta_0) r_i(u_\theta - \eta_i)| \right]$$

Expected penalty for errors in transmitted data

$$W(\tau) \triangleq \frac{1}{2} (\varphi(\tau) - \varphi_*)^2$$

Constraint penalty

$$\tau = \text{vec}[\theta, \eta_0, \eta_i]$$

Parameters

$$u_\theta \triangleq f(x, \theta)$$

Decision function

$$\eta_0, \eta_i$$

Thresholds

$$\varphi(\tau) \triangleq \mathbb{E} [\hat{I}_{\eta_0}(u_\theta)]$$

Transmission probability

$\hat{I}_{\eta_0}, \hat{I}_{\eta_i}, I_i \in \{0, 1\}$ indicate whether an observation z with features x is to be transmitted, classified as a positive, and is a true positive.

$I_e(z, \tau) \triangleq (1 - I_i(z)) \hat{I}_{\eta_i}(u_\theta) + I_i(z) (1 - \hat{I}_{\eta_i}(u_\theta))$ is the error indicator.

r_o, r_i are loss functions based on "distances" $u_\theta - \eta_0$ and $u_\theta - \eta_i$.

In the simplest case, $r_i(u_\theta - \eta_i) = u_\theta - \eta_i$.

DECISION RULES UPDATE

Stochastic
quasi-gradient
algorithm

$$\begin{aligned}\theta_{t+1} &= \theta_t - \gamma_{1,t} \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t} \hat{g}_\theta(\tilde{x}_{t+1}^\Psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t) \\ \eta_{0,t+1} &= \eta_{0,t} - \gamma_{3,t} \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\Psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t) \\ \eta_{1,t+1} &= \max \{ \eta_{0,t+1}, \eta_{1,t} - \gamma_{1,t} \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t) \}\end{aligned}$$
(a)

Quasi-gradient
of risk U

$$\begin{aligned}\hat{G}_\theta(\{z_i\}_{i=1}^n, \tau) &\triangleq \frac{1}{n} \sum_{i=1}^n [J(z_i, \tau) \nabla_\theta [r_o(u_i - \eta_0) r_i(u_i - \eta_1)]] \\ \hat{G}_{\eta_1}(\{z_i\}_{i=1}^n, \tau) &\triangleq -\frac{1}{n} \sum_{i=1}^n [J(z_i, \tau) r_o(u_i - \eta_0) \frac{\partial}{\partial u} r_i(u_i - \eta_1)]\end{aligned}$$

where $x_i = \chi(z_i)$, $u_i = f(x_i, \theta)$,

$$J(z, \tau) \triangleq \hat{I}_{\eta_0}(f(\chi(z), \theta)) (\hat{I}_{\eta_1}(f(\chi(z), \theta)) - I_1(z))$$
(b)

Quasi-gradient
of constraint

$$\begin{aligned}\hat{g}_\theta(\tilde{x}^\Psi, \tilde{x}^\Psi, \tau, \varepsilon) &\triangleq (\hat{\psi}(\tilde{x}^\Psi, \tau) - \varphi_*) \hat{\psi}_\theta(\tilde{x}^\Psi, \tau, \varepsilon) \\ \hat{g}_{\eta_0}(\tilde{x}^\Psi, \tilde{x}^\Psi, \tau, \varepsilon) &\triangleq (\hat{\psi}(\tilde{x}^\Psi, \tau) - \varphi_*) \hat{\psi}_{\eta_0}(\tilde{x}^\Psi, \tau, \varepsilon) \\ \hat{\psi}(\{x_i\}_{i=1}^n, \tau) &\triangleq \frac{1}{n} \sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta)) \\ \hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) &\triangleq \frac{1}{\varepsilon n} \sum_{i=1}^n [I_{S(\tau, \varepsilon)}(f(x_i, \theta)) \nabla_\theta f(x_i, \theta)] \\ \hat{\psi}_{\eta_0}(\{x_i\}_{i=1}^n, \tau, \varepsilon) &\triangleq -\hat{\psi}_0 = \text{const} < 0\end{aligned}$$
(c)

DECISION RULES UPDATE

$$\theta_{t+1} = \theta_t - \gamma_{1,t} \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t} \hat{g}_\theta(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t} \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max \{ \eta_{0,t+1}, \eta_{1,t} - \gamma_{1,t} \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t) \}$$

(a)

(a) Stochastic quasi-gradient algorithm

- (b) Quasi-gradient \hat{G} of the risk U
- (c) Quasi-gradient \hat{g} of the constraint term W

DECISION RULES UPDATE

$$\theta_{t+1} = \theta_t - \gamma_1, - \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_2, t \hat{g}_\theta(\tilde{x}_{t+1}^\psi, \tilde{\bar{x}}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_3, t \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\psi, \tilde{\bar{x}}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max \{ \eta_{0,t+1}, \eta_{1,t} - \gamma_1, t \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t) \} \quad (a)$$

(a) Stochastic quasi-gradient algorithm

- (b) Quasi-gradient \hat{G} of the risk U
- (c) Quasi-gradient \hat{g} of the constraint term W
- Updates are done in batches $\bar{z}_{t+1} = \{z_i\}$, with instrumental sub-batches $\tilde{x}_{t+1}^\psi, \tilde{\bar{x}}_{t+1}^\psi$ extracted from \bar{z}_{t+1} .

DECISION RULES UPDATE

$$\theta_{t+1} = \theta_t - \gamma_{1,t} \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t} \hat{g}_\theta(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t} \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max \{\eta_{0,t+1}, \eta_{1,t} - \gamma_{1,t} \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t)\}$$

(a)

(a) Stochastic quasi-gradient algorithm

- (b) Quasi-gradient \hat{G} of the risk U
- (c) Quasi-gradient \hat{g} of the constraint term W
- Updates are done in batches $\bar{z}_{t+1} = \{z_i\}$, with instrumental sub-batches $\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi$ extracted from \bar{z}_{t+1} .
- Threshold η_1 is clipped with η_0 for the bandwidth constraint to prevail over the classification risk.

DECISION RULES UPDATE

$$\theta_{t+1} = \theta_t - \gamma_{1,t} \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t} \hat{g}_\theta(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t} \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max \{ \eta_{0,t+1}, \eta_{1,t} - \gamma_{1,t} \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t) \}$$

(a)

(a) Stochastic quasi-gradient algorithm

- (b) Quasi-gradient \hat{G} of the risk U
- (c) Quasi-gradient \hat{g} of the constraint term W
- Updates are done in batches $\bar{z}_{t+1} = \{z_i\}$, with instrumental sub-batches $\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\psi$ extracted from \bar{z}_{t+1} .
- Threshold η_1 is clipped with η_0 for the bandwidth constraint to prevail over the classification risk.
- Different learning rates $\gamma_{1,t}, \gamma_{2,t}, \gamma_{3,t}$.

DECISION RULES UPDATE

$$\begin{aligned}\theta_{t+1} &= \theta_t - \gamma_{1,t} \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t} \hat{g}_\theta(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t) \\ \eta_{0,t+1} &= \eta_{0,t} - \gamma_{3,t} \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t) \\ \eta_{1,t+1} &= \max \{ \eta_{0,t+1}, \eta_{1,t} - \gamma_{1,t} \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t) \}\end{aligned}\quad (a)$$

$$\hat{G}_\theta(\{z_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n} \sum_{i=1}^n [J(z_i, \tau) \nabla_\theta [r_o(u_i - \eta_0) r_i(u_i - \eta_1)]]$$

$$\hat{G}_{\eta_1}(\{z_i\}_{i=1}^n, \tau) \triangleq -\frac{1}{n} \sum_{i=1}^n [J(z_i, \tau) r_o(u_i - \eta_0) \frac{d}{du} r_i(u_i - \eta_1)]$$

where $x_i = \chi(z_i)$, $u_i = f(x_i, \theta)$,

$$J(z, \tau) \triangleq \hat{I}_{\eta_0}(f(\chi(z), \theta)) (\hat{I}_{\eta_1}(f(\chi(z), \theta)) - I_1(z)) \quad (b)$$

(b) Quasi-gradient \hat{G} of the risk U

$$\mathbb{E}[\hat{G}(\bar{z}, \tau)] = G(\tau), \text{ the generalized gradient such that } U(\tau_{t+1}) - U(\tau_t) \leq \langle \tau_{t+1} - \tau_t, G(\tau_t) \rangle + O(\|\tau_{t+1} - \tau_t\|^2).$$

DECISION RULES UPDATE

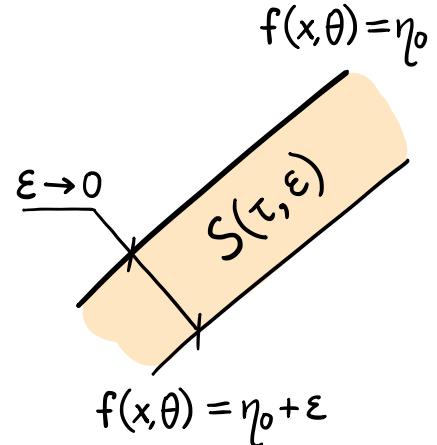
$$\hat{g}_\theta(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\psi}(\tilde{x}^\varphi, \tau) - \varphi_*) \hat{\psi}_\theta(\tilde{x}^\psi, \tau, \varepsilon)$$

$$\hat{g}_{\eta_0}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\psi}(\tilde{x}^\varphi, \tau) - \varphi_*) \hat{\psi}_\eta(\tilde{x}^\psi, \tau, \varepsilon)$$

$$\hat{\psi}(\{x_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n} \sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta))$$

$$\hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq \frac{1}{\varepsilon n} \sum_{i=1}^n [I_{S(\tau, \varepsilon)}(f(x_i, \theta)) \nabla_\theta f(x_i, \theta)]$$

$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\hat{\psi}_0 = \text{const} < 0$$



(c)

Quasi-gradient \hat{g} of the constraint term W

- Evolution of the integral in the gradient of W

$$\nabla_\tau W(\tau) = (\psi(\tau) - \varphi_*) \psi(\tau), \quad \text{where}$$

$$\psi(\tau) \triangleq \nabla_\tau \mathbb{E}[\hat{I}_{\eta_0}(u_\theta)] = \int_{\{f(x, \theta) = \eta_0\}} \frac{\nabla_\tau [f(x, \theta) - \eta_0]}{\| \nabla_x f(x, \theta) \|} p(x) d\sigma$$

$$= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{S(\tau, \varepsilon)} \nabla_\tau [f(x, \theta) - \eta_0] p(x) dx$$

volume



surface



volume

$$S(\tau, \varepsilon) \triangleq \{x : \eta_0 < f(x, \theta) \leq \eta_0 + \varepsilon\}, \quad I_{S(\tau, \varepsilon)}(u) \triangleq \mathbb{1}[u \in S(\tau, \varepsilon)]$$

DECISION RULES UPDATE

$$\theta_{t+1} = \theta_t - \gamma_{1,t} \hat{G}_\theta(\bar{x}_{t+1}, \tau_t) - \gamma_{2,t} \hat{g}_\theta(\tilde{x}_{t+1}^\Psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t)$$

$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t} \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\Psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t)$$

$$\eta_{1,t+1} = \max \{ \eta_{0,t+1}, \eta_{1,t} - \gamma_{1,t} \hat{G}_\eta(\bar{x}_{t+1}, \tau_t) \}$$

(a)

$$\hat{g}_\theta(\tilde{x}^\Psi, \tilde{x}^\Psi, \tau, \varepsilon) \triangleq (\hat{\psi}(\tilde{x}^\Psi, \tau) - \varphi_*) \hat{\psi}_\theta(\tilde{x}^\Psi, \tau, \varepsilon)$$

$$\hat{g}_{\eta_0}(\tilde{x}^\Psi, \tilde{x}^\Psi, \tau, \varepsilon) \triangleq (\hat{\psi}(\tilde{x}^\Psi, \tau) - \varphi_*) \hat{\psi}_\eta(\tilde{x}^\Psi, \tau, \varepsilon)$$

$$\hat{\psi}(\{x_i\}_{i=1}^n, \tau) \triangleq \frac{1}{n} \sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta))$$

$$\hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq \frac{1}{\varepsilon n} \sum_{i=1}^n [I_{S(\tau, \varepsilon)}(f(x_i, \theta)) \nabla_\theta f(x_i, \theta)]$$

$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\hat{\psi}_0 = \text{const} < 0$$

(c)

- Instrumental samples are used to make a stochastic estimator for the product of φ and ψ in $\nabla_\tau W(\tau)$:

$$\begin{aligned} \mathbb{E}[\hat{g}(\tilde{x}^\Psi, \tilde{x}^\Psi, \tau, \varepsilon)] &= \mathbb{E}[(\hat{\psi}(\tilde{x}^\Psi, \tau) - \varphi_*) \hat{\psi}(\tilde{x}^\Psi, \tau, \varepsilon)] + O(\varepsilon) \\ &= \mathbb{E}[\hat{\psi}(\tilde{x}^\Psi, \tau) - \varphi_*] \mathbb{E}[\hat{\psi}(\tilde{x}^\Psi, \tau, \varepsilon)] + O(\varepsilon) \end{aligned}$$

DECISION RULES UPDATE

$$\begin{aligned}\theta_{t+1} &= \theta_t - \gamma_{1,t} \hat{G}_\theta(\bar{z}_{t+1}, \tau_t) - \gamma_{2,t} \hat{g}_\theta(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t) \\ \eta_{0,t+1} &= \eta_{0,t} - \gamma_{3,t} \hat{g}_{\eta_0}(\tilde{x}_{t+1}^\psi, \tilde{x}_{t+1}^\Psi, \tau_t, \varepsilon_t) \\ \eta_{1,t+1} &= \max \{ \eta_{0,t+1}, \eta_{1,t} - \gamma_{1,t} \hat{G}_{\eta_1}(\bar{z}_{t+1}, \tau_t) \}\end{aligned}\quad (a)$$

$$\begin{aligned}\hat{g}_\theta(\tilde{x}^\psi, \tilde{x}^\Psi, \tau, \varepsilon) &\triangleq (\hat{\psi}(\tilde{x}^\psi, \tau) - \varphi_*) \hat{\psi}_\theta(\tilde{x}^\psi, \tau, \varepsilon) \\ \hat{g}_{\eta_0}(\tilde{x}^\psi, \tilde{x}^\Psi, \tau, \varepsilon) &\triangleq (\hat{\psi}(\tilde{x}^\psi, \tau) - \varphi_*) \hat{\psi}_\eta(\tilde{x}^\psi, \tau, \varepsilon) \\ \hat{\psi}(\{x_i\}_{i=1}^n, \tau) &\triangleq \frac{1}{n} \sum_{i=1}^n \hat{I}_{\eta_0}(f(x_i, \theta)) \\ \hat{\psi}_\theta(\{x_i\}_{i=1}^n, \tau, \varepsilon) &\triangleq \frac{1}{\varepsilon n} \sum_{i=1}^n [I_{S(\tau, \varepsilon)}(f(x_i, \theta)) \nabla_\theta f(x_i, \theta)] \\ \hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) &\triangleq -\hat{\psi}_0 = \text{const} < 0\end{aligned}\quad (c)$$

- As in the above, the gradient of the integral leads to:

$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n, \tau, \varepsilon) \triangleq -\frac{1}{\varepsilon n} \sum_{i=1}^n I_{S(\tau, \varepsilon)}(f(x_i, \theta)) \leq 0$$

Yet, in the algorithm, we use a **constant estimator** for promptness in the updates of η_0 , as higher adaptivity is of more value than accuracy for adjusting to the bandwidth constraint.

DECISION RULES UPDATE

Theorem (Algorithm convergence) Let

- batches \bar{z}_t be i.i.d.;
- sub-batches \tilde{x}_t^ψ and \tilde{x}_t^θ be mutually independent;
- $r_0(u)$, $r_1(u)$, and $f(x, \theta)$ be continuously differentiable;
- $\text{sgn}(r_i(u)) = \text{sgn}(u)$, $i \in \{0, 1\}$; $r_i(u) \leq r_i(v)$ for all $0 \leq u < v$;
- p.d.f. $p(x)$ of observations in the feature space be continuous and have compact support;
- $\gamma_{1,t}/\gamma_{2,t}$ and ε_t be monotonically decreasing
 $\sum \gamma_{i,t}^2 < \infty$, $i \in \{1, 2, 3\}$; $\sum \gamma_{j,t} \varepsilon_t < \infty$, $j \in \{1, 2\}$;
 $\sum \gamma_{3,t} \gamma_{1,t}/\gamma_{2,t} < \infty$; $\sum \gamma_{3,t} = \infty$; $\sum \gamma_{1,t}^2/\gamma_{2,t} = \infty$.

Then, for $\zeta_t = \gamma_{2,t}/\gamma_{1,t}$,

- $\frac{1}{\zeta_t} U(\tau_t) + W(\tau_t) \xrightarrow[t \rightarrow \infty]{\text{a.s.}} V_*$, $\mathbb{E}[V_*] < \infty$;
- $\lim_{t \rightarrow \infty} |\psi(\tau_t) - \psi_*| = 0$;
- $\lim_{t \rightarrow \infty} G_{\eta_1}(\tau_t) = 0$;
- $\lim_{t \rightarrow \infty} G_\theta(\tau_t) + \zeta_t \nabla_\theta W(\tau_t) = 0$.

RANDOMIZED EDGE-ASSISTED ON-SENSOR INFORMATION SELECTION FOR BANDWIDTH-CONSTRAINED SYSTEMS

Igor Burago
iburago@uci.edu

Marco Levorato
levorato@uci.edu

Intelligent and Autonomous Systems Lab
<https://iasl.ics.uci.edu>

Paper
<http://ib.fyi/acssc18>

