

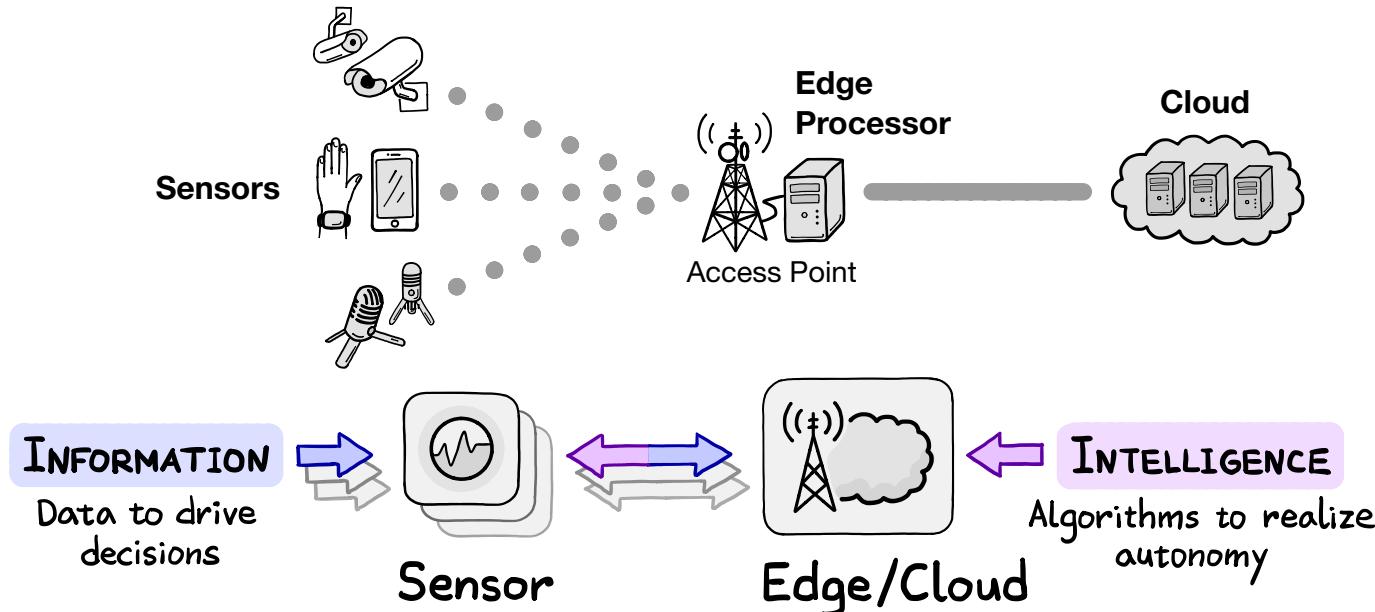
CLOUD-ASSISTED ON-SENSOR OBSERVATION CLASSIFICATION IN LATENCY-IMPEDED IOT SYSTEMS

Igor Burago Marco Levorato

Department of Computer Science
University of California, Irvine



INTELLIGENT IoT IN PRINCIPLE

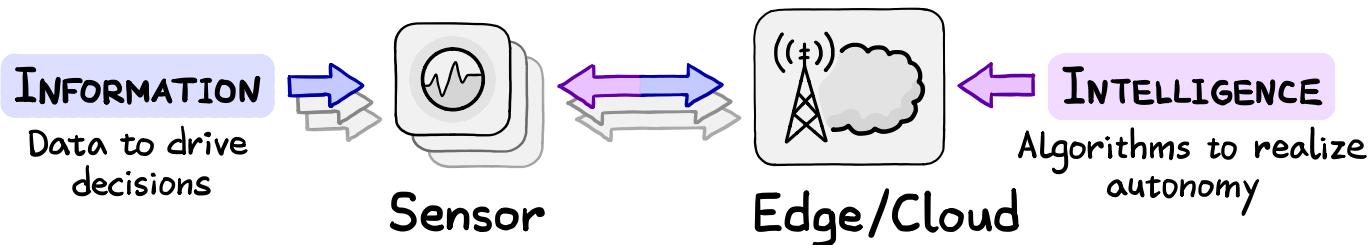


- Observes the environment
- Selectively informs the edge
- Adjusts its own operation

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

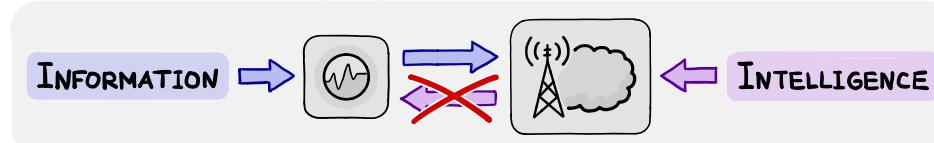
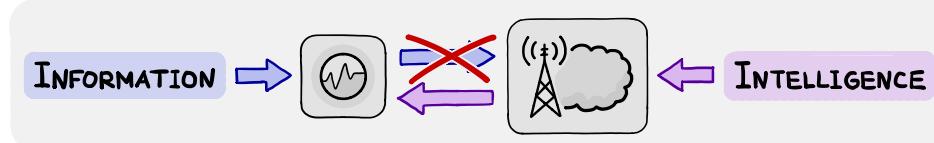
Goal: Realize the alliance between
information completeness and best intelligence

INTELLIGENT IoT CHALLENGE



System is constrained by:

- Sensor High-quality decision cannot be made locally
AND
- Channel with bottleneck in:
 - Bandwidth Edge/Cloud cannot see all observations in full
 - OR
 - Latency Decision turnaround is too slow for reactive sensor adjustments



APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

Strategies vary in the amount of:

Decision
Quality

APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

Strategies vary in the amount of:

Decision
Quality On-Sensor
 Computation

APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

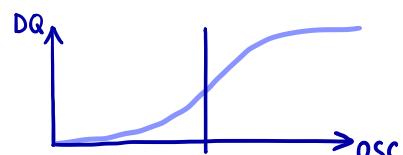
Strategies vary in the amount of:

Decision Quality	On-Sensor Computation	Sensor-Cloud Communication
------------------	-----------------------	----------------------------

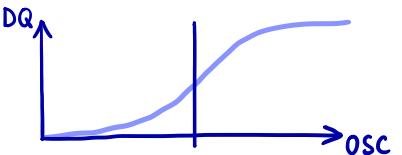
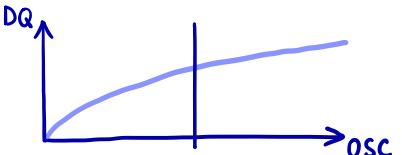
APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

	Decision / On-Sensor Quality / Computation	Sensor-Cloud Communication
• Offload all decisions to cloud	BEST	NONE

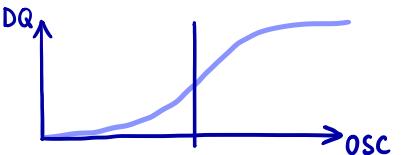
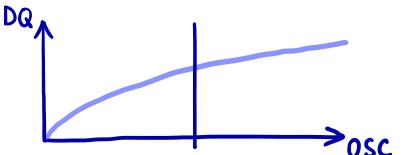
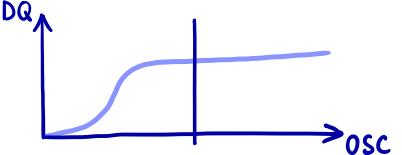
APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

	Decision / On-Sensor Quality / Computation	Sensor-Cloud Communication
• Offload all decisions to cloud	BEST	NONE
• Make all decisions on sensor		HIGH Low

APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

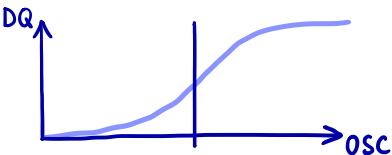
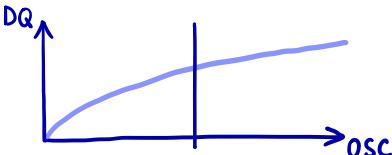
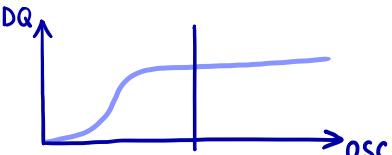
	Decision / On-Sensor Quality / Computation	Sensor-Cloud Communication
• Offload all decisions to cloud	BEST	NONE
• Make all decisions on sensor		HIGH Low
• Compress observations and offload all decisions to cloud		MEDIUM

APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

	Decision / On-Sensor Quality / Computation	Sensor-Cloud Communication
• Offload all decisions to cloud	BEST	NONE
• Make all decisions on sensor		HIGH Low
• Compress observations and offload all decisions to cloud		MEDIUM
• Selectively transmit observations to cloud for decisions*		Low

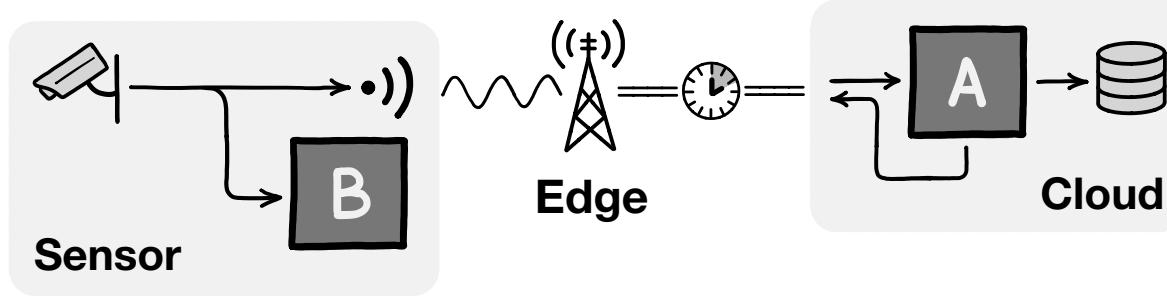
* I. Burago and M. Levorato, "Randomized Edge-Assisted On-Sensor Information Selection for Bandwidth-Constrained Systems", Asilomar Conference on Signals, Systems, and Computers, 2018.

APPROACHES TO COMPUTATION-COMMUNICATION TRADEOFF

	Decision Quality / On-Sensor Computation	Sensor-Cloud Communication	Decision Turnaround
• Offload all decisions to cloud	BEST	NONE	HIGH DELAYED
• Make all decisions on sensor		Low	NONE
• Compress observations and offload all decisions to cloud		MEDIUM	DELAYED
• Selectively transmit observations to cloud for decisions*		Low	DELAYED

* I. Burago and M. Levorato, "Randomized Edge-Assisted On-Sensor Information Selection for Bandwidth-Constrained Systems", Asilomar Conference on Signals, Systems, and Computers, 2018.

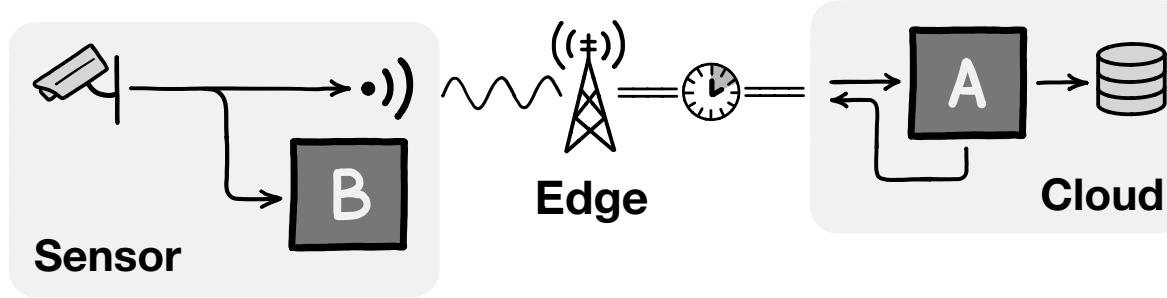
SENSOR-CLOUD COOPERATION MODEL



Model assumptions:

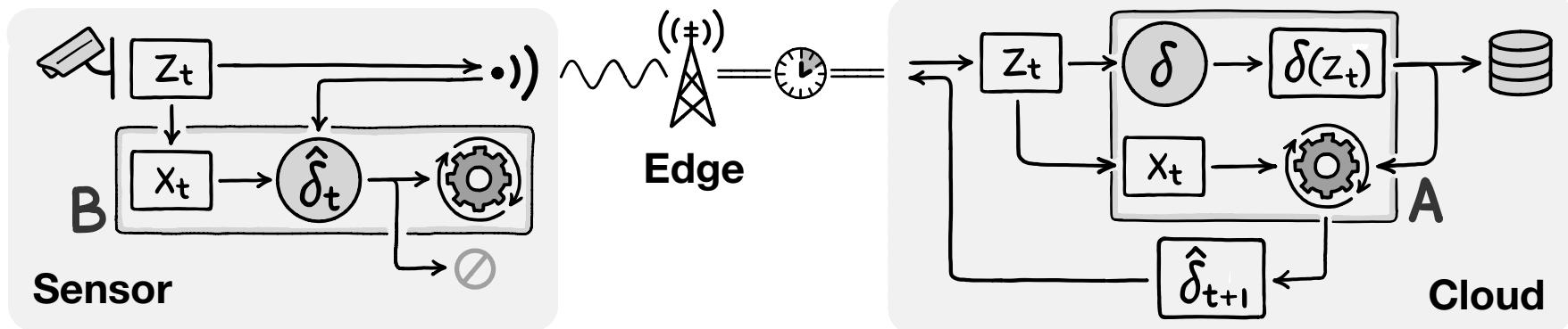
- Observations fall into two classes:
 - of interest to the sensor (positives)
 - and not (negatives)
 - Sensor adjusts to its environment by reacting to positives
 - Cloud can run the optimal classifier δ ; sensor can not
 - Sensor-cloud link has sufficient bandwidth but unacceptable latency for sensor adaptivity
- B**
- Sensor runs a low-complexity classifier, supplied by cloud, to make local decisions
- A**
- Cloud, via supervision of its classifier, trains the sensor's

SENSOR-CLOUD COOPERATION MODEL



Factors at play

- Delayed channel
- CPU, battery, etc.
- No tolerance for missing decisions
- Computation constraints
- Auxiliary decision $\hat{\delta}$ to learn
- Complexity of $\hat{\delta}$
- Need to react to environment
- Cost of error in feature space
- Expected risk V to minimize



SENSOR-CLOUD COOPERATION MODEL

Problem Formulation

For the on-sensor decision rule $\hat{\delta}: f(x, \theta) \geq \mu$

$$V(\tau) \triangleq \mathbb{E}[I_E(z, \tau) |r(u_\theta - \mu)|] \longrightarrow \min_{\tau} \text{ s.t. } \psi(\tau) \leq \psi_*$$

Expected penalty for errors

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

Parameters

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

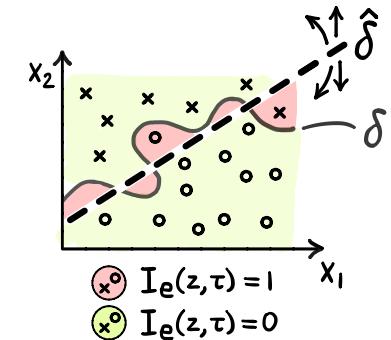
Decision function

$$\mu$$

Threshold

$$\psi(\tau) \triangleq \mathbb{E}[\hat{I}_\mu^{(1)}(u_\theta)]$$

Probability of positives



$$\psi_* > \pi_* \triangleq \mathbb{E}[I_1(z)]$$

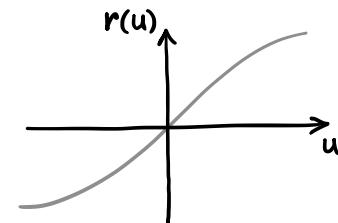
Potential positive classifiability

$\hat{I}_\mu^{(1)}$, $I_1 \in \{0, 1\}$ indicate whether an observation z with features x is classified as a positive and is a true positive, respectively.

$I_E(z, \tau) \triangleq (1 - I_1(z)) \hat{I}_\mu^{(1)}(u_\theta) + I_1(z) (1 - \hat{I}_\mu^{(1)}(u_\theta))$ is the error indicator.

r is a loss function based on the "distance" $u_\theta - \mu$.

In the simplest case, $r(u_\theta - \mu) = u_\theta - \mu$.



CHOICE OF CRITERION

Error probability ($r(\cdot) \triangleq 1$):

$$\mathbb{E}[I_E(z, \tau)] \longrightarrow \min_{\tau}$$

Challenging to build a gradient estimator.

For instance, let $F(\tau) \triangleq \int_{\{x : f(x, \tau) > 0\}} p(x) dx$.

$$\text{Then } \nabla F(\tau) = - \int_{\{x : f(x, \tau) > 0\}} \operatorname{div}_x [\Lambda(x, \tau) p(x)] dx, \quad (\text{A})$$

$$\Lambda(x, \tau) \triangleq \frac{\nabla_\tau f(x, \tau) (\nabla_x f(x, \tau))^\top}{\|\nabla_x f(x, \tau)\|}$$

$$\text{or } \nabla F(\tau) = \int_{\{x : f(x, \tau) = 0\}} \frac{\nabla_\tau f(x, \tau)}{\|\nabla_x f(x, \tau)\|} p(x) d\sigma. \quad (\text{B})$$

(A) requires estimating $p(x)$ for $\nabla p(x)$. \leftarrow Practically impossible

(B) requires estimating a surface integral in X . \leftarrow Slow

CHOICE OF CRITERION

Error probability ($r(\cdot) \triangleq 1$):

$$\mathbb{E}[I_E(z, \tau)] \longrightarrow \min_{\tau}$$

Challenging to build a gradient estimator.

For instance, let $F(\tau) \triangleq \int_{\{x : f(x, \tau) > 0\}} p(x) dx$.

$$\text{Then } \nabla F(\tau) = - \int_{\{x : f(x, \tau) > 0\}} \operatorname{div}_x [\Lambda(x, \tau) p(x)] dx, \quad (\text{A})$$

$$\Lambda(x, \tau) \triangleq \frac{\nabla_\tau f(x, \tau) (\nabla_x f(x, \tau))^\top}{\|\nabla_x f(x, \tau)\|}$$

$$\text{or } \nabla F(\tau) = \int_{\{x : f(x, \tau) = 0\}} \frac{\nabla_\tau f(x, \tau)}{\|\nabla_x f(x, \tau)\|} p(x) d\sigma. \quad (\text{B})$$

(A) requires estimating $p(x)$ for $\nabla p(x)$. \leftarrow Practically impossible

(B) requires estimating a surface integral in X . \leftarrow Slow

Expected risk (nontrivial $r(\cdot)$): $V(\tau) \triangleq \mathbb{E}[I_E(z, \tau) | r(u_\theta - \mu)|] \longrightarrow \min_{\tau}$

Practical stochastic gradient is possible when

$I_E(z, \tau)$ and $r(\cdot)$ concord to allow $I_E(z, \tau) \cdot |r(\cdot)|$ to be differentiable

In $V(\tau)$: $r(u_\theta - \mu) > 0$ when $I_E(z, \tau) = 1$

$r(u_\theta - \mu) \leq 0$ when $I_E(z, \tau) = 0$

DECISION RULE UPDATE

Algorithm

$$\tau_{t+1} = \mathcal{H}(\tau_t - \gamma_t \hat{G}(z_{t+1}, \tau_t)),$$

where $\hat{G}(z, \tau) \triangleq \text{vec}[J(z, \tau) \nabla_\theta f(x, \theta), -J(z, \tau)]$
is a stochastic quasi-gradient of $V(\tau)$;

$J(z, \tau) \triangleq \begin{cases} +1, & \text{for false positive;} \\ 0, & \text{for correct;} \\ -1, & \text{for false negative;} \end{cases}$
is the signed error indicator;

γ_t is learning rate; and

$\mathcal{H}(\tau)$ is a transformation encapsulating the a priori knowledge on parameter localization (or $\mathcal{H}(\tau) \equiv \tau$).

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies

$$|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u'';$$

- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau;$

- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{\text{a.s.}}{=} V_*, \quad \mathbb{E}V_* < \infty;$
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{\text{a.s.}}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right].$

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

1) a sample of i.i.d. observations $\{z_t\}$;

2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

3) a feature-extraction function $\chi(z)$, s.t.

the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies

$$|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u'';$$

4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau$;

5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{\text{a.s.}}{=} V_*, \quad \mathbb{E}V_* < \infty$;

- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:

$$\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{\text{a.s.}}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right].$$

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$\begin{aligned} |f(x, \theta'') - f(x, \theta')| &\leq C_f(x) \|\theta'' - \theta'\|, & \mathbb{E}[(C_f(x))^k] &\leq C_{f,k}, \quad k > 1, \\ |\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| &\leq C_{\nabla f}(x) \|\theta'' - \theta'\|, & \mathbb{E}[C_{\nabla f}(x)] &\leq C_{\nabla f}; \end{aligned}$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies
 $|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u''$;
- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau$;
- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{a.s.}{=} V_*, \quad \mathbb{E}V_* < \infty$;
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{a.s.}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right]$.

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies

$$|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u'';$$

- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau;$
- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{\text{a.s.}}{=} V_*, \quad \mathbb{E}V_* < \infty;$
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{\text{a.s.}}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right].$

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies

$$|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u'';$$

- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau;$

- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{\text{a.s.}}{=} V_*, \quad \mathbb{E}V_* < \infty;$
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{\text{a.s.}}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right].$

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies
 $|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u''$;
- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau$;
- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{\text{a.s.}}{=} V_*, \quad \mathbb{E}V_* < \infty$;
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{\text{a.s.}}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right]$.

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies

$$|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u'';$$

- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau;$

- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

E.g., if $\gamma_t = Ct^{-\alpha}$, $C, \alpha > 0$,
then $\frac{k+1}{2k} < \alpha \leq 1$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{a.s.}{=} V_*, \quad \mathbb{E}V_* < \infty;$
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{a.s.}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right].$

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies

$$|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u'';$$

- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau;$

- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{\text{a.s.}}{=} V_*, \quad \mathbb{E}V_* < \infty;$
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{\text{a.s.}}{=} 0, \quad G(\tau) \triangleq \text{vec}\left[\mathbb{E}[J(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[J(z, \tau)]\right].$

DECISION RULE UPDATE GUARANTEES

Theorem 1 (Algorithm convergence) Given

- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) a continuously differentiable decision function $f(x, \theta)$, s.t. $\forall x, \theta', \theta''$

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[(C_f(x))^k] \leq C_{f,k}, \quad k > 1,$$

$$|\nabla_\theta f(x, \theta'') - \nabla_\theta f(x, \theta')| \leq C_{\nabla f}(x) \|\theta'' - \theta'\|, \quad \mathbb{E}[C_{\nabla f}(x)] \leq C_{\nabla f};$$

- 3) a feature-extraction function $\chi(z)$, s.t.
the c.d.f. F_{u_θ} of the random variable $u_\theta \triangleq f(\chi(z), \theta)$ satisfies
 $|F_{u_\theta}(u'') - F_{u_\theta}(u')| \leq C_F |u'' - u'|, \quad \forall \theta, u', u''$;
- 4) a mapping $\mathcal{H}(\tau)$, s.t. $V(\mathcal{H}(\tau)) \leq V(\tau) + C_H \|\mathcal{H}(\tau) - \tau\|^2 \quad \forall \tau$;
- 5) a sequence of learning rates $\gamma_t > 0$, s.t.

$$\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^{2\lambda} < \infty, \quad \lambda \triangleq \frac{k}{k+1};$$

the algorithm:

- converges in criterion: $\lim_{t \rightarrow \infty} V(\tau_t) \stackrel{a.s.}{=} V_*, \quad \mathbb{E}V_* < \infty$;
- achieves the necessary condition of extremum for $V(\tau)$ on a subsequence:
 $\liminf_{t \rightarrow \infty} \|G(\tau_t)\| \stackrel{a.s.}{=} 0, \quad G(\tau) \triangleq \text{vec} \left[\mathbb{E}[\mathbf{J}(z, \tau) \nabla_\theta f(x, \theta)], -\mathbb{E}[\mathbf{J}(z, \tau)] \right]$.

DECISION RULE UPDATE GUARANTEES

Theorem 2 (Constraint satisfaction)

Under the assumptions of Theorem 1, if

1) condition 2) of Theorem 1 holds for $\kappa = \infty$, i.e.,

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \text{ where } \mathbb{E}[(C_f(x))^\kappa] \leq C_{f,\kappa} \quad \forall \kappa;$$

2) p.d.f. $p(x)$ of observations in the feature space that is continuous and has compact support;

3) $\mathcal{H}(\tau + \gamma \Delta) - \tau = \gamma \Psi(\tau) \Delta + O(\gamma^2)$ for some bounded matrix, $\|\Psi(\tau)\| \leq C_\Psi \quad \forall \tau$;

then, on the trajectory of the algorithm,

$\psi(\tau_t)$ converges to $\pi_* < \psi_*$ in mean square and w.p. 1.

DECISION RULE UPDATE GUARANTEES

Theorem 2 (Constraint satisfaction)

Under the assumptions of Theorem 1, if

1) condition 2) of Theorem 1 holds for $\kappa = \infty$, i.e.,

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \text{ where } \mathbb{E}[(C_f(x))^\kappa] \leq C_{f,\kappa} \quad \forall \kappa;$$

2) p.d.f. $p(x)$ of observations in the feature space that is continuous and has compact support;

3) $\mathcal{H}(\tau + \gamma \Delta) - \tau = \gamma \Psi(\tau) \Delta + O(\gamma^2)$ for some bounded matrix, $\|\Psi(\tau)\| \leq C_\Psi \quad \forall \tau$;

then, on the trajectory of the algorithm,

$\psi(\tau_t)$ converges to $\pi_* < \psi_*$ in mean square and w.p. 1.

DECISION RULE UPDATE GUARANTEES

Theorem 2 (Constraint satisfaction)

Under the assumptions of Theorem 1, if

1) condition 2) of Theorem 1 holds for $\kappa = \infty$, i.e.,

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \text{ where } \mathbb{E}[(C_f(x))^\kappa] \leq C_{f,\kappa} \quad \forall \kappa;$$

2) p.d.f. $p(x)$ of observations in the feature space that is continuous and has compact support;

3) $\mathcal{H}(\tau + \gamma \Delta) - \tau = \gamma \Psi(\tau) \Delta + O(\gamma^2)$ for some bounded matrix, $\|\Psi(\tau)\| \leq C_\Psi \quad \forall \tau$;

then, on the trajectory of the algorithm,

$\psi(\tau_t)$ converges to $\pi_* < \psi_*$ in mean square and w.p. 1.

DECISION RULE UPDATE GUARANTEES

Theorem 2 (Constraint satisfaction)

Under the assumptions of Theorem 1, if

1) condition 2) of Theorem 1 holds for $\kappa = \infty$, i.e.,

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \text{ where } \mathbb{E}[(C_f(x))^\kappa] \leq C_{f,\kappa} \quad \forall \kappa;$$

2) p.d.f. $p(x)$ of observations in the feature space that is continuous and has compact support;

3) $\mathcal{H}(\tau + \gamma \Delta) - \tau = \gamma \Psi(\tau) \Delta + O(\gamma^2)$ for some bounded matrix, $\|\Psi(\tau)\| \leq C_\Psi \quad \forall \tau$;

then, on the trajectory of the algorithm,

$\psi(\tau_t)$ converges to $\pi_* < \psi_*$ in mean square and w.p. 1.

DECISION RULE UPDATE GUARANTEES

Theorem 2 (Constraint satisfaction)

Under the assumptions of Theorem 1, if

1) condition 2) of Theorem 1 holds for $\kappa = \infty$, i.e.,

$$|f(x, \theta'') - f(x, \theta')| \leq C_f(x) \|\theta'' - \theta'\|, \text{ where } \mathbb{E}[(C_f(x))^\kappa] \leq C_{f,\kappa} \quad \forall \kappa;$$

2) p.d.f. $p(x)$ of observations in the feature space that is continuous and has compact support;

3) $\mathcal{H}(\tau + \gamma \Delta) - \tau = \gamma \Psi(\tau) \Delta + O(\gamma^2)$ for some bounded matrix, $\|\Psi(\tau)\| \leq C_\Psi \quad \forall \tau$;

then, on the trajectory of the algorithm,

$\psi(\tau_t)$ converges to $\pi_* < \psi_*$ in mean square and w.p. 1.

DECISION RULE UPDATE GUARANTEES

Algorithm (Hyperplane separator):

$$\hat{\theta}_{t+1} = \theta_t - \gamma_t J(z_{t+1}, \tau_t) (\mathbf{I} - \theta_{t+1} \theta_{t+1}^\top) x_{t+1},$$

$$\mu_{t+1} = \mu_t + \gamma_t J(z_{t+1}, \tau_t)$$

$$\tau_{t+1} = \text{vec} \left[\hat{\theta}_{t+1} / \|\hat{\theta}_{t+1}\|, \mu_{t+1} \right]$$

In other words,

$$f(x, \theta) \triangleq \frac{\langle x, \theta \rangle}{\|\theta\|},$$

$$\mathcal{H}(\tau) \triangleq \text{vec} \left[\theta / \|\theta\|, \mu \right].$$

DECISION RULE UPDATE GUARANTEES

Algorithm (Hyperplane separator):

$$\hat{\theta}_{t+1} = \theta_t - \gamma_t J(z_{t+1}, \tau_t) (\mathbf{I} - \theta_{t+1} \theta_{t+1}^\top) x_{t+1},$$

$$\mu_{t+1} = \mu_t + \gamma_t J(z_{t+1}, \tau_t)$$

$$\tau_{t+1} = \text{vec} \left[\hat{\theta}_{t+1} / \|\hat{\theta}_{t+1}\|, \mu_{t+1} \right]$$

In other words,

$$f(x, \theta) \triangleq \frac{\langle x, \theta \rangle}{\|\theta\|},$$

$$\mathcal{H}(\tau) \triangleq \text{vec} \left[\theta / \|\theta\|, \mu \right].$$

Theorem 3 (Hyperplane separator) Given

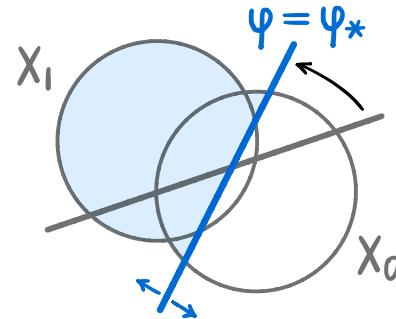
- 1) a sample of i.i.d. observations $\{z_t\}$;
- 2) p.d.f. $p(x)$ of observations in the feature space that is continuous and have compact support;
- 3) $\sum_{t=1}^{\infty} \gamma_t = \infty, \quad \sum_{t=1}^{\infty} \gamma_t^2 < \infty$;

the algorithm converges in criterion and achieves

$$\lim_{t \rightarrow \infty} \left\| \mathbb{E}_z [J(z_{t+1}, \tau_t) (\mathbf{I} - \theta_{t+1} \theta_{t+1}^\top) x_{t+1}] \right\| \stackrel{\text{a.s.}}{=} 0,$$

$$\lim_{t \rightarrow \infty} \mathbb{E}_z [J(z_{t+1}, \tau_t)] \stackrel{\text{a.s.}}{=} 0,$$

i.e., $\psi(\tau_t)$ converges to $\pi_1 < \psi_*$ w.p. 1.



CLOUD-ASSISTED ON-SENSOR OBSERVATION CLASSIFICATION IN LATENCY-IMPEDED IOT SYSTEMS

Igor Burago

iburago@uci.edu

Marco Levorato

levorato@uci.edu

Intelligent and Autonomous Systems Lab

iasl.ics.uci.edu

Papers

- ib.fyi/isit19
- ib.fyi/acsc18

