**1. Intelligent IoT in Principle**

- **Information**: Data to drive decisions
- **Intelligence**: Algorithms to realize autonomy

Goal: Realize the alliance between information completeness and best intelligence

Schematic view into the space of computation-communication tradeoffs

**Model assumptions**
- **Observations fall into two classes**: of interest to the system (positives) and not (negatives)
- **Edge can run the optimal classifier** $\delta$; sensor can not
- **Sensor runs a low-complexity classifier** supplied by edge
- **Edge**, via supervision of its classifier, trains the sensor’s
- **Sensor uses its classifier to spare bandwidth by dropping some negatives, and to make local decisions**

**Factors at play**
- **CPU, battery, etc.**
- **Computation constraints**
- **Complexity of $\delta$ and $\delta$**
- **Bandwidth constraint**
- **Cost of error in feature space**
- **Need to react to environment**
- **Transmission probability $\psi = \psi_0$**
- **Expected risk $U$ to minimize**

**Problem**: $U(c) \rightarrow \min$, s.t. $W(c) = 0$

- $U(c) = \mathbb{E} [\mathbb{I}_p (u_0) I_2 (z(x)) f(z(x) - \mu_0 + \eta_0) + \psi (x)]$ Expected penalty for errors in transmitted data
- $z(x) = \text{vec} [\theta_0 \beta_x, \eta_x]$
- $\psi = \mathbb{E} [\mathbb{I}_p (u_0) f (z(x) - \mu_0 + \eta_0)]$
- $\eta_0$, $\eta_1$ are loss functions based on “distances” $u_0 - \eta_0$ and $u_0 - \eta_1$
- In the simplest case, $\eta_1 (u_0 - \eta_1) = u_0 - \eta_1$

**2. Sensor-Edge Cooperation Model**

**Edge/Cloud**
- Observations transferred to the server
- Processing allocated on the server
- Effect induced on the quality of decisions

**Decision Quality**
- Offload all decisions to edge
- Offload all decisions to sensor
- Sample observations to transmit
- Transmit obvious positives
- Transmit all in compressed form
- Optimal adaptive compromise

**3. Decision Rules Update**

$\theta_{t+1} = \theta_0 - \eta_{t} \mathbb{E} [\mathbb{I}_p (u_0) I_2 (z(x)) (1 - I_0 (z(x)) - I_0 (1 - I_0 (z(x))))]$

**4. Algorithm Specifics Overview**

**Stochastic quasi-gradient algorithm**
- Updates done in batches $E = \{z_i\}$, with instrumental sub-batches $\eta_{it} = \eta_{t}$ extracted from $E$
- Threshold $\eta_{it}$ is clipped with $\eta_0$ for the bandwidth constraint to prevent over the classification risk

**Quasi-gradient $\delta$ of the risk $U$**

$E [\mathbb{I}_p (u_0) f(z(x)) + \psi (x)]$, where

**Instrumental samples are used to make a stochastic estimator** for the product of $\eta$ and $\psi$ in $V(c)$:

$E \left[ \mathbb{I}_p (u_0) f(z(x)) + \psi (x) \right] = \mathbb{E} \left[ \mathbb{I}_p (u_0) f(z(x)) + \psi (x) \right] + O(c)$

- **As in the above, the gradients of the integral leads to**: $\hat{\theta}_{t+1} = \theta_0 - \frac{\psi (x)}{\sum \mathbb{I}_p (u_0) f(z(x))} \leq 0$

**Theorem (Algorithm convergence)**

- Batches $E_i$ be i.i.d.
- Sub-batches $X_i$ and $X_i$ be mutually independent
- $\mathbb{E} [u_i] = \mathbb{E} [u_i]$, and $\mathbb{E} [u_i] = \mathbb{E} [u_i]$ be continuously differentiable
- $\mathbb{E} [u_i] = \mathbb{E} [u_i]$, and $\mathbb{E} [u_i]$ be $\mathbb{E} [u_i]$ for all $0 \leq u \leq V$
- p.d.f. $p(x)$ of observations in the feature space be continuous and have compact support
- $\mathbb{E} [u_i] = \mathbb{E} [u_i]$, and $\mathbb{E} [u_i]$ be monotonically decreasing
- $\sum \mathbb{E} [u_i] \leq \sum \mathbb{E} [u_i]$ $\Rightarrow$ $\mathbb{E} [u_i] \leq \mathbb{E} [u_i]$
- $\mathbb{E} [u_i] \leq \mathbb{E} [u_i]$, and $\mathbb{E} [u_i] = \mathbb{E} [u_i]$
- Then, for $E_i = E_i / N_i$

- $\mathbb{E} [u_i] = \mathbb{E} [u_i]$, $\mathbb{E} [u_i] = \mathbb{E} [u_i]$, $\mathbb{E} [u_i]$ $\Rightarrow$ $\mathbb{E} [u_i] = \mathbb{E} [u_i]$
- $\lim t \rightarrow \infty \mathbb{I}_p (u_0) f(z(x)) = 0$
- $\lim t \rightarrow 0 \mathbb{I}_p (u_0) f(z(x)) = 0$
- $\lim t \rightarrow 0 \mathbb{I}_p (u_0) f(z(x)) = 0$
- $\lim t \rightarrow 0 \mathbb{I}_p (u_0) f(z(x)) = 0$