On Active sampling of controlled experiments for QoE modeling

Muhammad Jawad Khokhar

Nawfal Abbasi Saber

Thiery Spetebroot

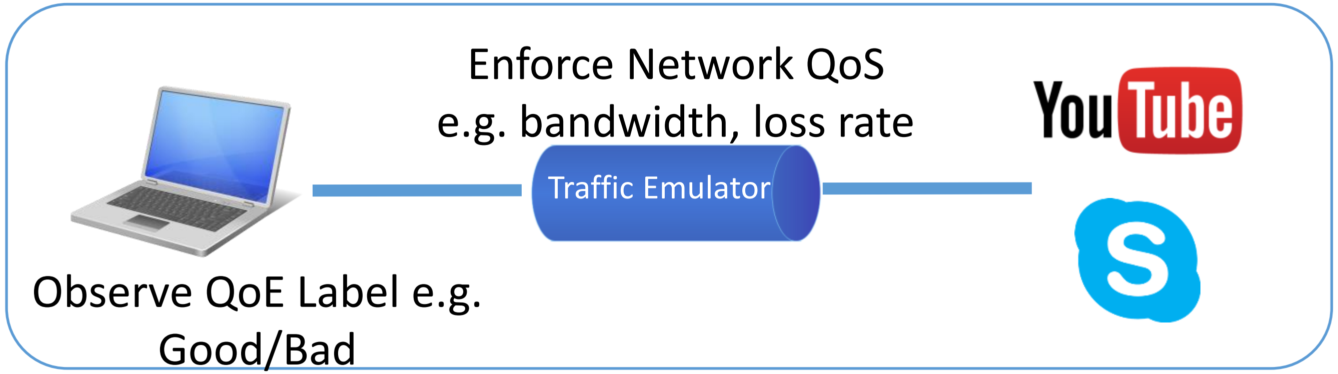
Chadi Barakat

ACM SIGCOMM 2017 2nd Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE 2017), August 21, 2017

INVENTEURS DU MONDE NUMÉRIQUE

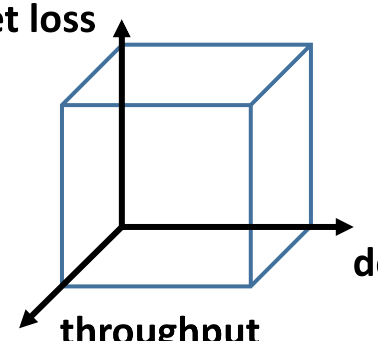
A common approach for QoE modeling: ML and controlled experimentation

A common approach for QoE modeling: ML and controlled experimentation



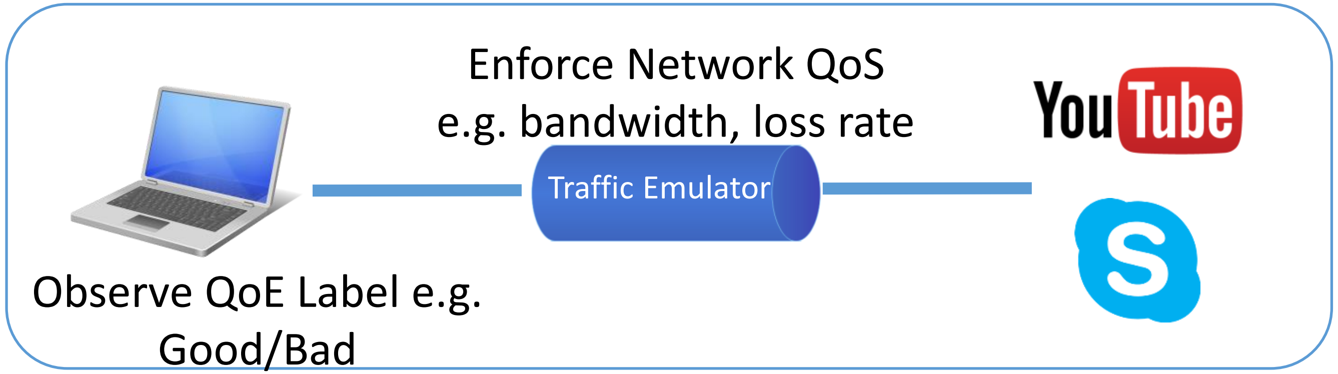
A common approach for QoE modeling: ML and controlled experimentation

Experimental space



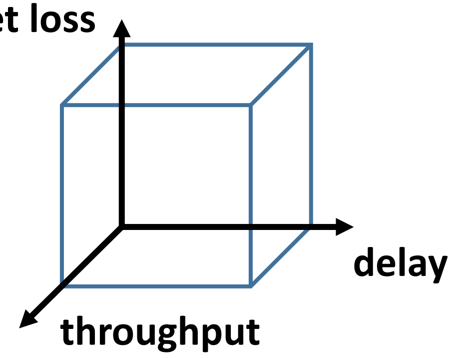
Set of network configurations

delay	loss	throughput	QoE
100ms	10%	10Mbps	?
…	…	…	?
300ms	0.1%	5Mbps	?



A common approach for QoE modeling: ML and controlled experimentation

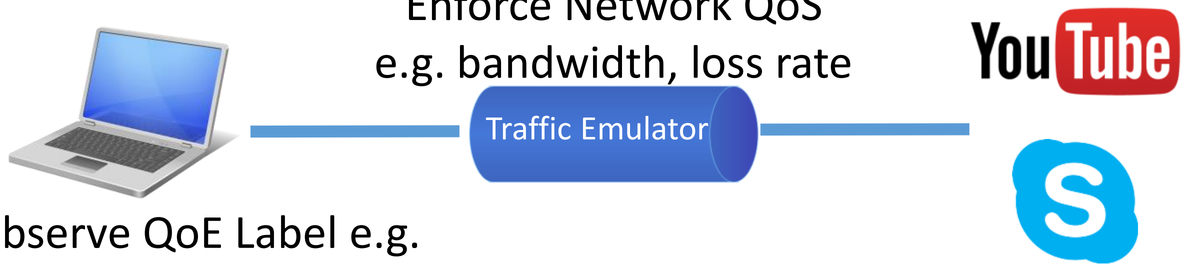
Experimental space



Set of network configurations

delay	loss	throughput	QoE
100ms	10%	10Mbps	?
…	…	…	?
300ms	0.1%	5Mbps	?

Enforce Network QoS
e.g. bandwidth, loss rate



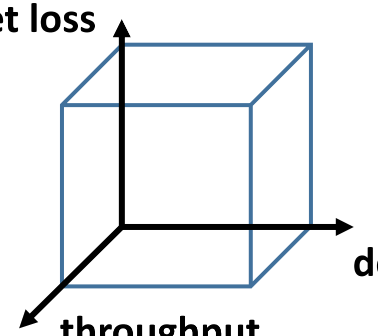
Observe QoE Label e.g. Good/Bad

Training Dataset

delay	loss	throughput	QoE
100ms	10%	10Mbps	Good
…	…	…	Bad
300ms	0.1%	5Mbps	Good

A common approach for QoE modeling: ML and controlled experimentation

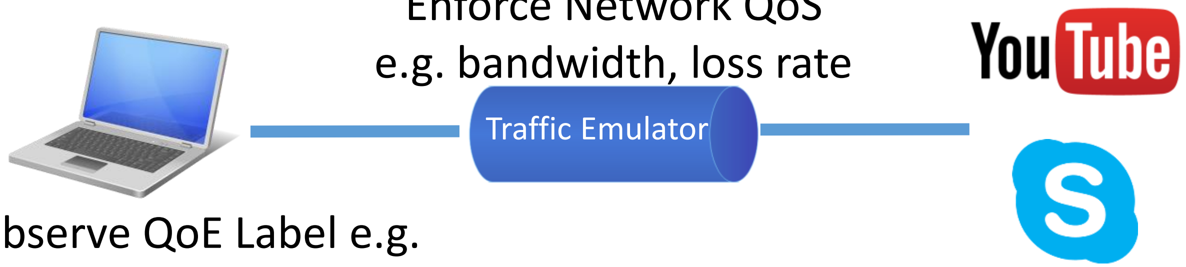
Experimental space



Set of network configurations

delay	loss	throughput	QoE
100ms	10%	10Mbps	?
…	…	…	?
300ms	0.1%	5Mbps	?

Enforce Network QoS
e.g. bandwidth, loss rate



Observe QoE Label e.g. Good/Bad

Training Dataset

delay	loss	throughput	QoE
100ms	10%	10Mbps	Good
…	…	…	Bad
300ms	0.1%	5Mbps	Good

Network QoS → **ML Model** → **Application QoE**

INVENTEURS DU MONDE NUMÉRIQUE

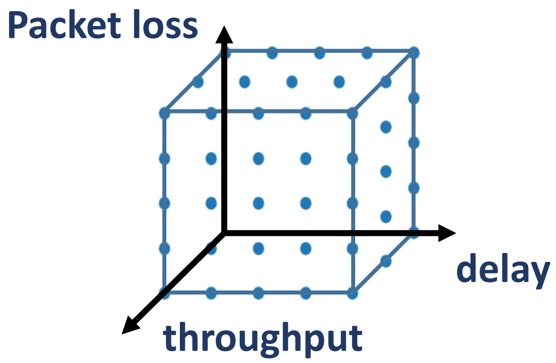
The Problem Formulation

Accurate QoE modeling requires building large training sets

The Problem Formulation

Accurate QoE modeling requires building large training sets

A Conventional approach:
UNIFORM SAMPLING



INVENTEURS DU MONDE NUMÉRIQUE

The Problem Formulation

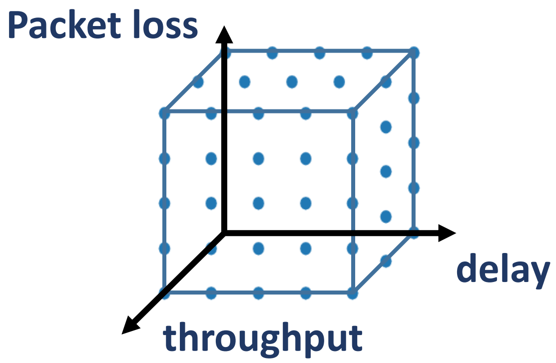
Accurate QoE modeling requires building large training sets

A challenge in controlled experimentation:

High training cost

A Conventional approach:

UNIFORM SAMPLING



e.g. with 2 min for each experiment,
10000 experiments = 20000 minutes
= 14 days of continuous experimentation

INVENTEURS DU MONDE NUMÉRIQUE

The Problem Formulation

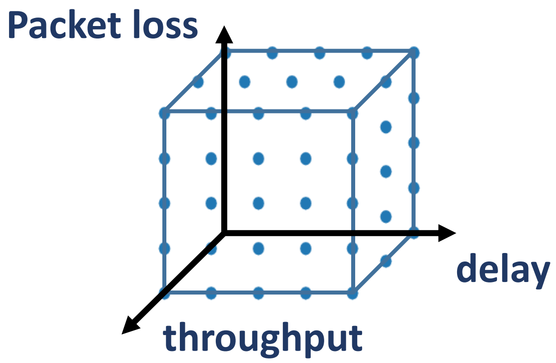
Accurate QoE modeling requires building large training sets

A challenge in controlled experimentation:

High training cost

Can we reduce **training cost** while not impacting modeling accuracy?

A Conventional approach:
UNIFORM SAMPLING



e.g. with 2 min for each experiment,
10000 experiments = 20000 minutes
= 14 days of continuous experimentation

Uniform sampling amplifies the training cost with little improvement in accuracy.

The Problem Formulation

Accurate QoE modeling requires building large training sets

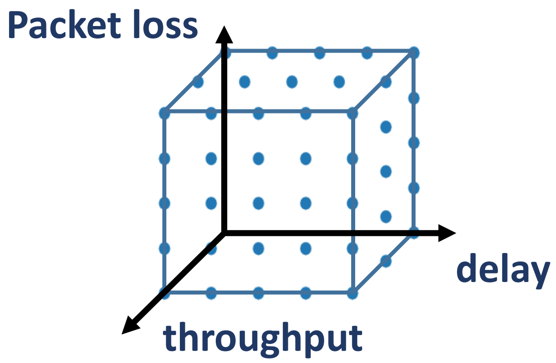
A challenge in controlled experimentation:

High training cost

Can we reduce **training cost** while not impacting modeling accuracy?

Proposed solution: **Active Learning**

A Conventional approach:
UNIFORM SAMPLING



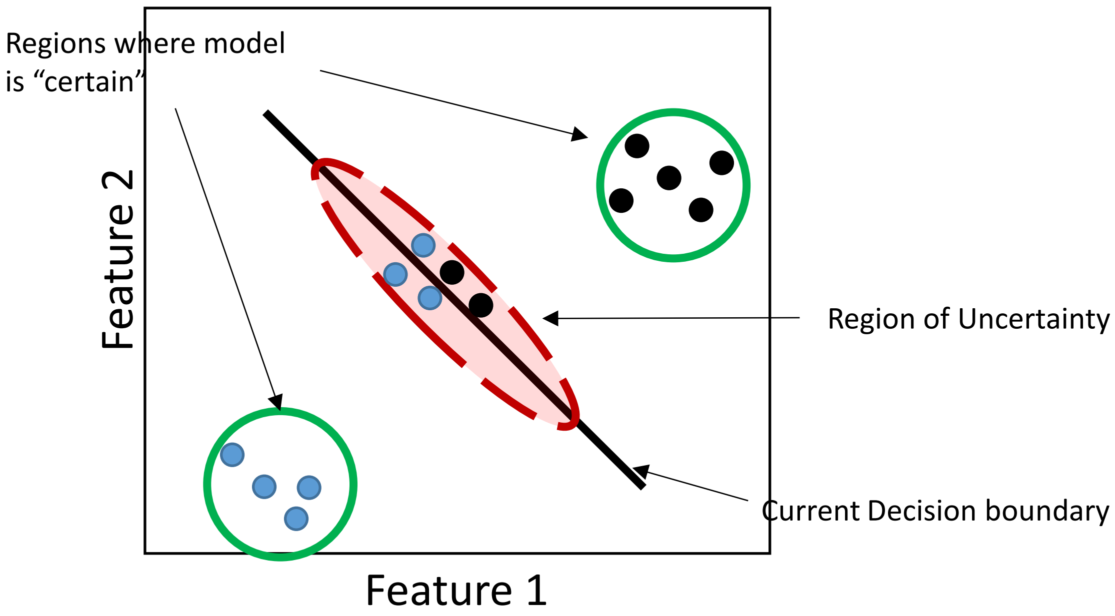
e.g. with 2 min for each experiment,
10000 experiments = 20000 minutes
= 14 days of continuous experimentation

Uniform sampling amplifies the training cost with little improvement in accuracy.

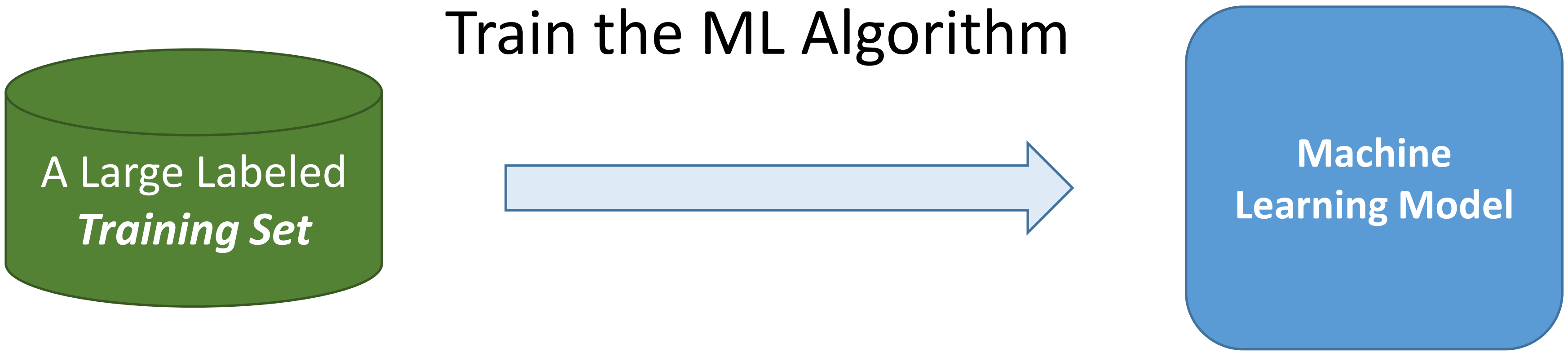
Experiment in the **most useful** regions that bring **maximum gain** in the **accuracy** of the model.

INVENTEURS DU MONDE NUMÉRIQUE

Useful regions in the experimental space: Regions of uncertainty



Conventional Supervised Machine Learning

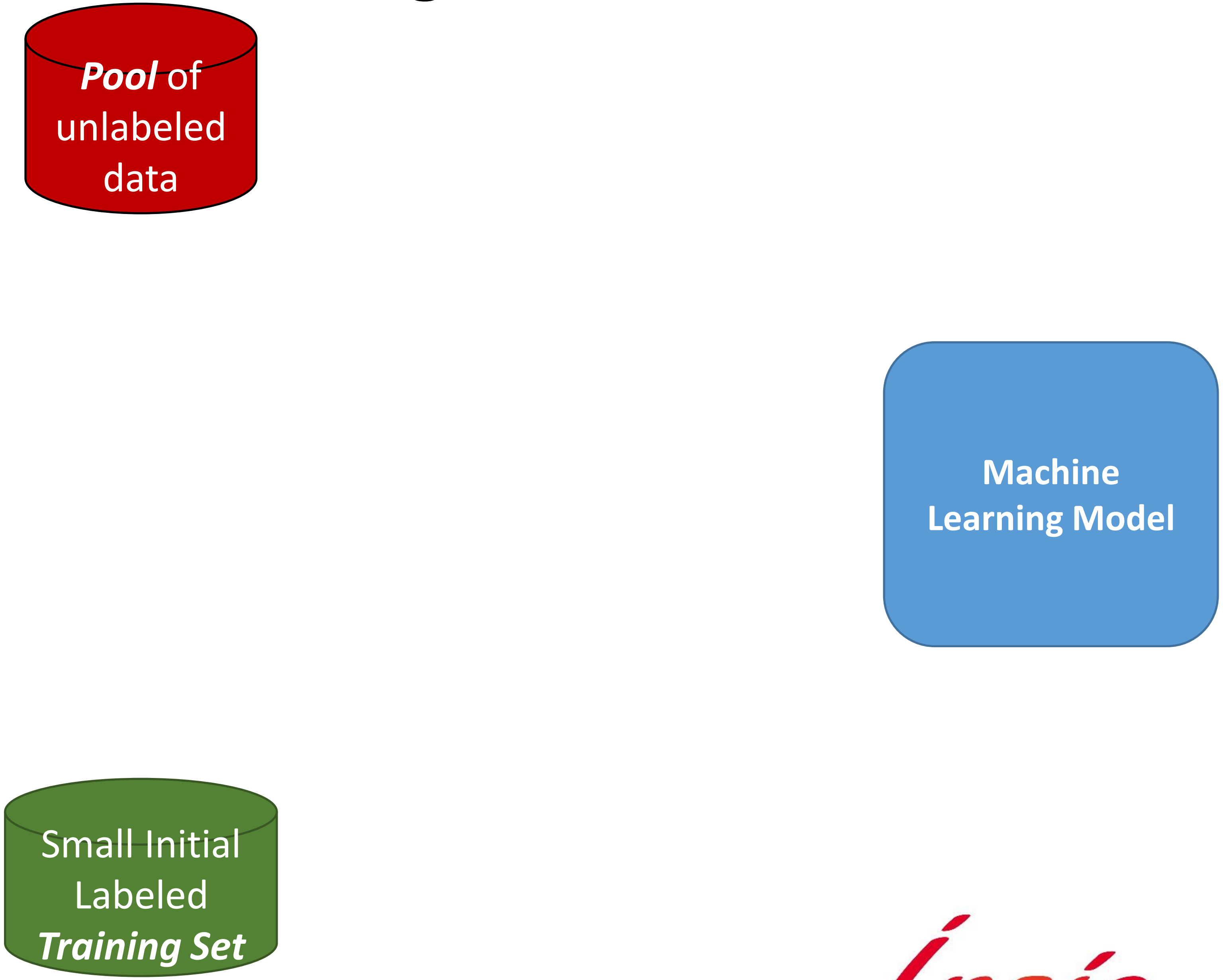


delay	loss	throughput	QoE
...	...	...	Good
...	...	...	Bad
...	...	...	Good

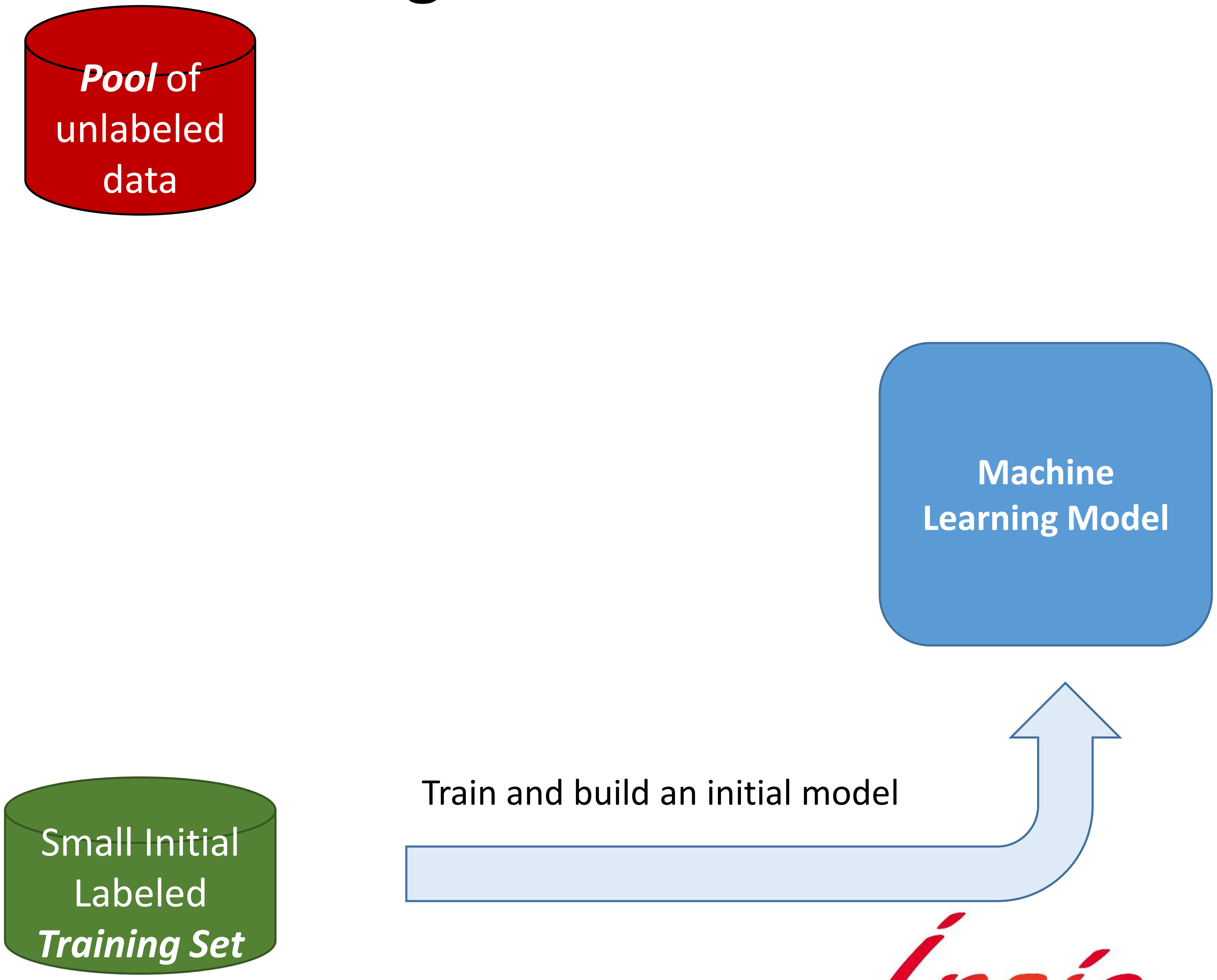
Features Label

INVENTEURS DU MONDE NUMÉRIQUE

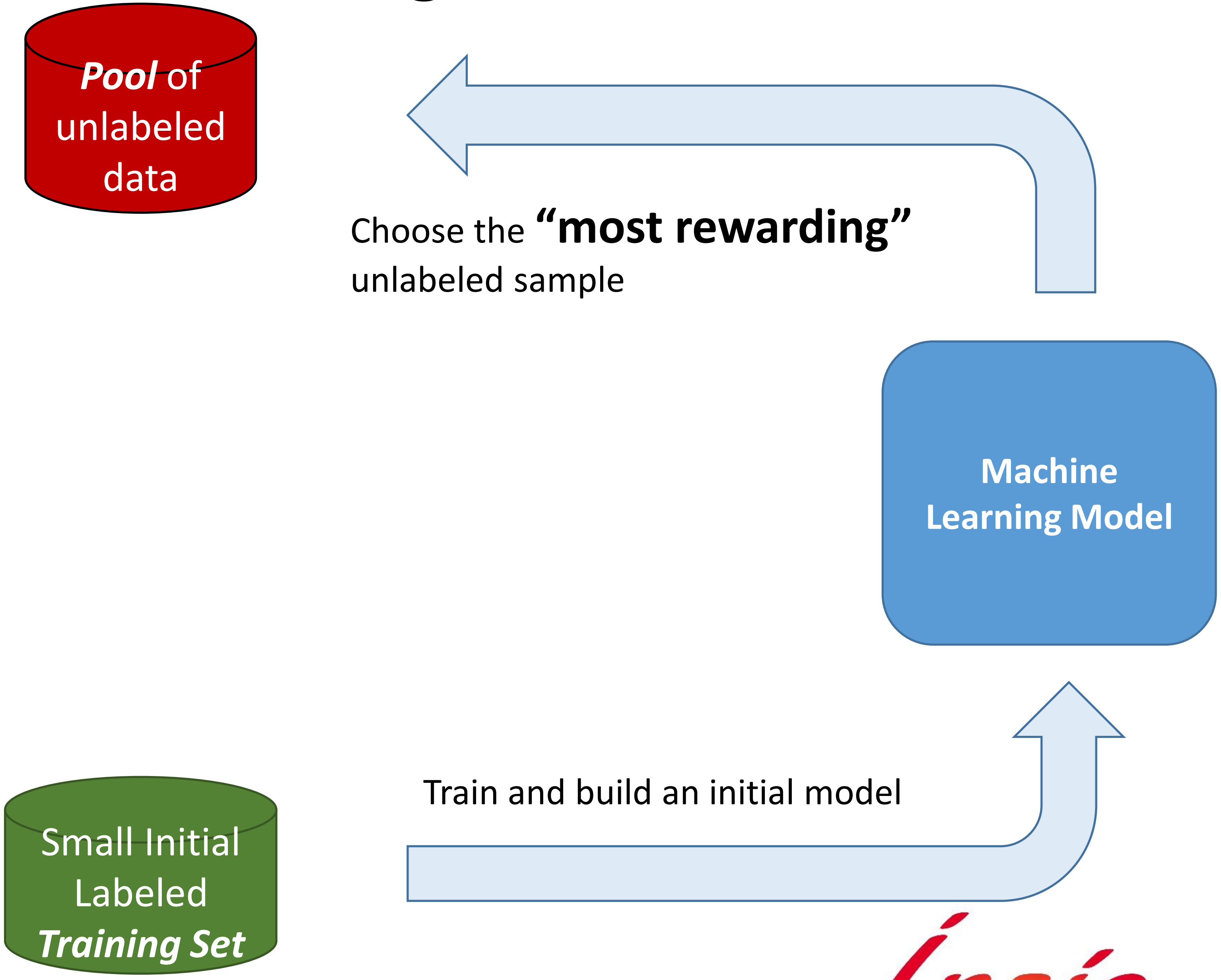
Active Learning



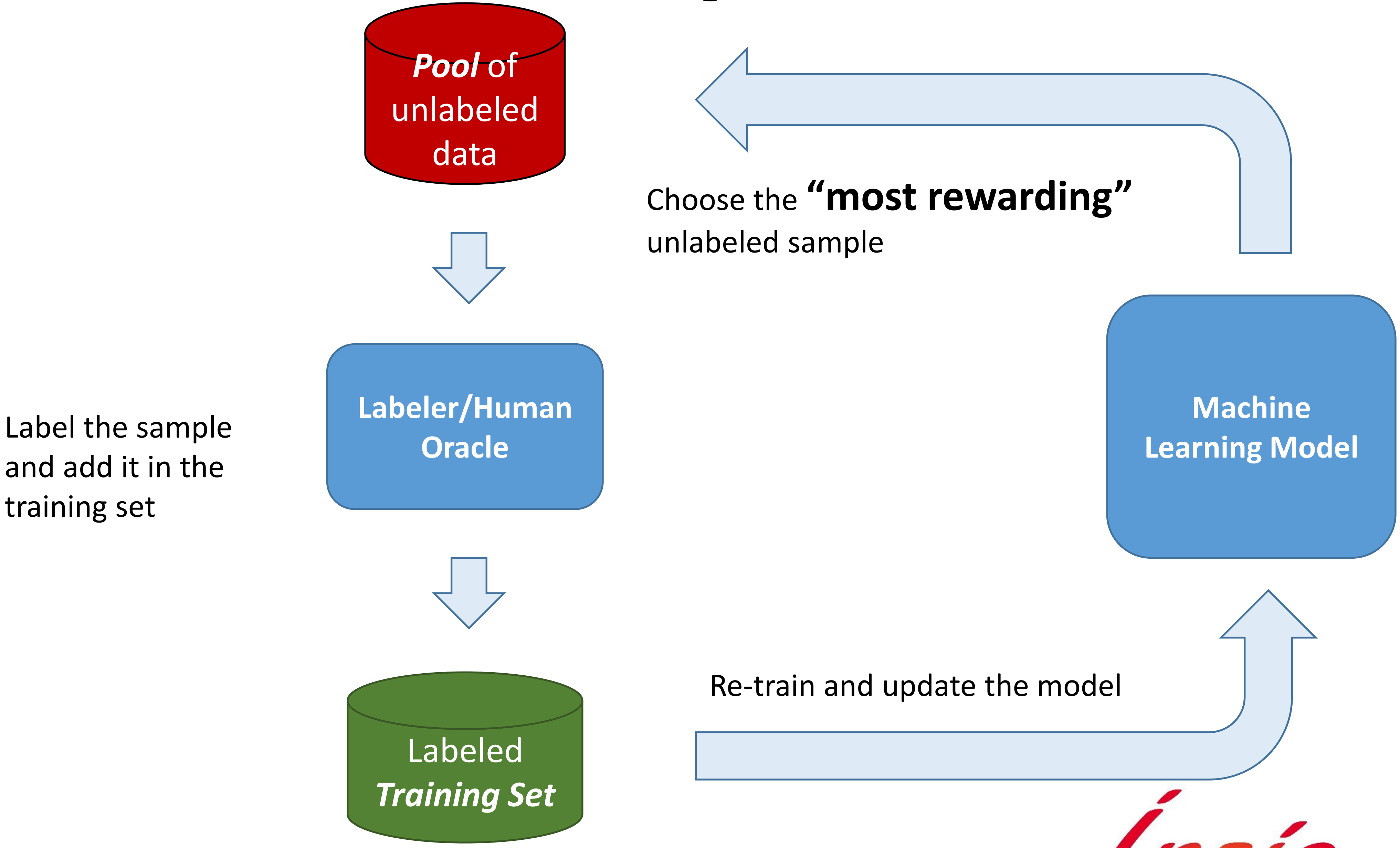
Active Learning



Active Learning

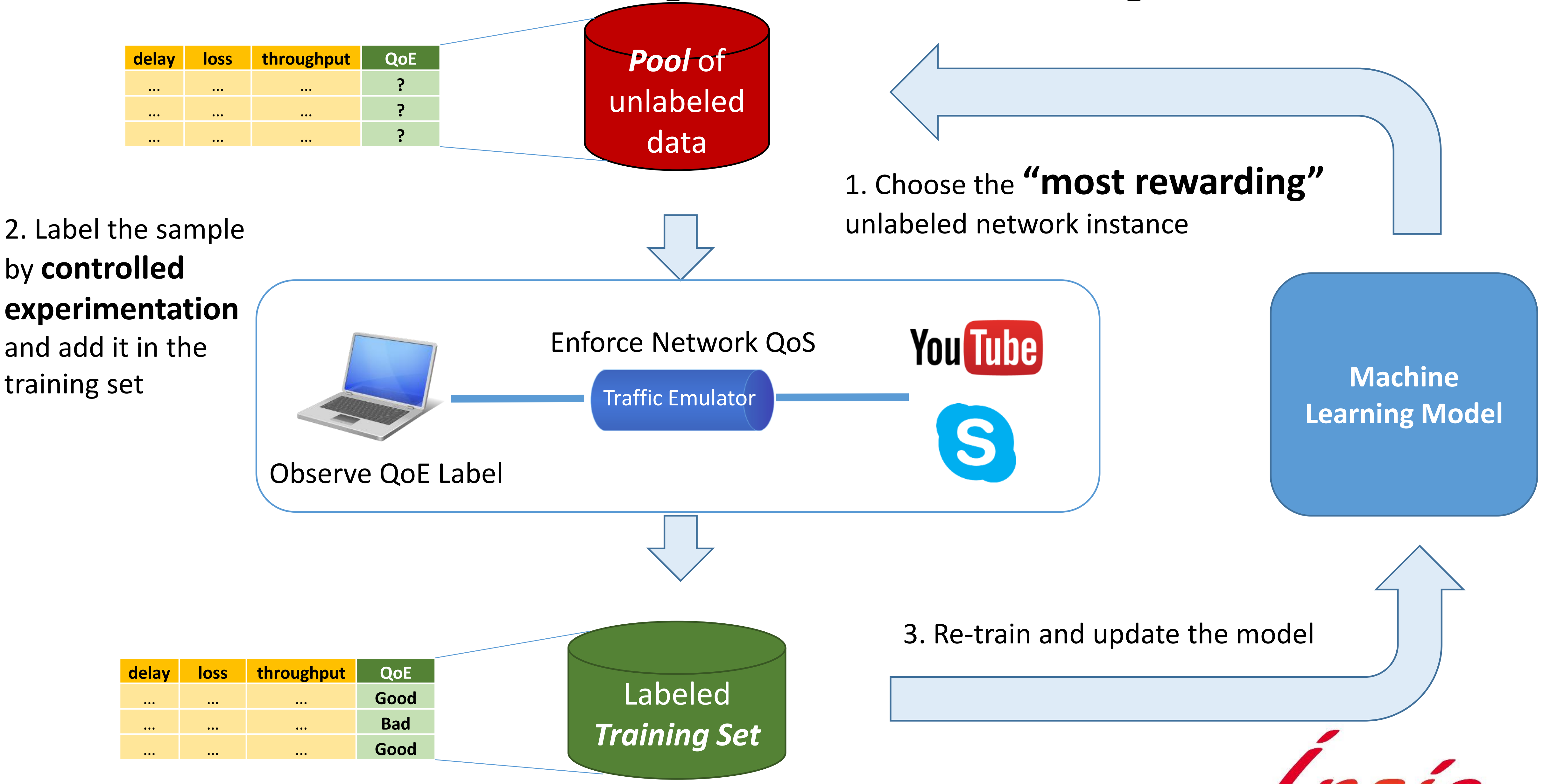


Active Learning



Active Learning for QoE Modeling

delay	loss	throughput	QoE
…	…	…	?
…	…	…	?
…	…	…	?



delay	loss	throughput	QoE
…	…	…	Good
…	…	…	Bad
…	…	…	Good

INVENTEURS DU MONDE NUMÉRIQUE

Method of choosing the most rewarding sample from the Pool

Model's classification probability for each output label/class

Pool of Unlabeled Data

	$P(\hat{y}^{(1)})$	$P(\hat{y}^{(2)})$	$P(\hat{y}^{(3)})$	$P(\hat{y}^{(4)})$	$P(\hat{y}^{(5)})$
$x^{(1)}$	0.1	0.35	0.1	0.2	0.25
$x^{(2)}$					
…					
$x^{(p)}$					

$$\sum_i P(\hat{y}^{(i)}) = 1$$

Least Confident: $\text{argmin}_x P(\hat{y})$

Minimal Margin: $\text{argmin}_x [P(\hat{y}_1) - P(\hat{y}_2)]$

Maximum Entropy: $\text{argmax}_x - \sum_y P(y) \log P(y)$

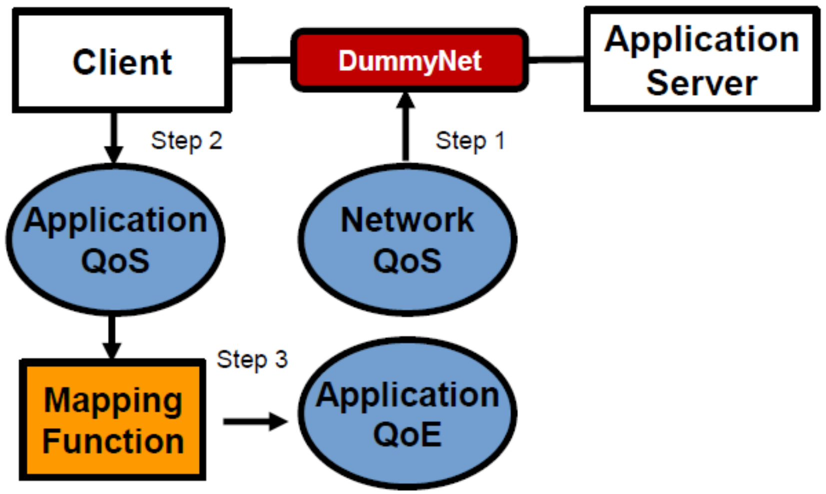
Overall methodology

$\mathcal{P}$ = Pool of unlabeled instances $\{x^{(p)}\}_{p=1}^{P}$
$\mathcal{T}$ = Training set of labeled instances $\{\langle x, y\rangle^{(t)}\}_{t=1}^{T}$
Θ = QoE Model e.g. a Decision Tree
Φ = Utility measure of Uncertainty e.g. Max Entropy
Initialize $\mathcal{T}$
for $i = 1, 2, \ldots$ **do**
    Θ = **train**($\mathcal{T}$)
    select $x^* \in \mathcal{P}$, as per Φ
    experiment using x^* to obtain label y^*
    add $\langle x^*, y^*\rangle$ to $\mathcal{T}$
    remove x^* from $\mathcal{P}$
end for

The YouTube use case

- **Network QoS features:**
  1. RTT
  2. Loss Rate
  3. Download Throughput/Bandwidth
- **Application QoS features:**
  1. Initial Join time
  2. Total duration of the stalling events

Dataset Labeling process



Mapping Function: QoE definition for YouTube Video

Binary Classification:

$$QoE_{binary} = \begin{cases} 0 \ - Bad\ (if\ video\ stalls) \\ 1 \ - Good\ (if\ video\ does\ not\ stall) \end{cases}$$

0 - Bad	1 - Good

Multiclass Classification:

$$QoE_{multi} = \alpha e^{-\beta t} + 1 \quad (\alpha = 4, \beta = 0.0347)$$

where t is the total buffering time and factors α and β are computed according to below assumptions for best and worst case scenarios:

1. Best case: QoE is maximum of 5 for zero buffering time;
2. Worst case: QoE is 1.5 for buffering of 50% of the total duration of the video

1 - Poor	2 - Bad	3 - Fair	4 - Good	5 - Excellent

INVENTEURS DU MONDE NUMÉRIQUE

The experimental space for dataset collection

- **Instances Pool:**
  - RTT = 0 – 5000 ms
  - Loss Rate = 0 – 25 %
  - Throughput = 0 – 10 Mbps

The experimental space for dataset collection

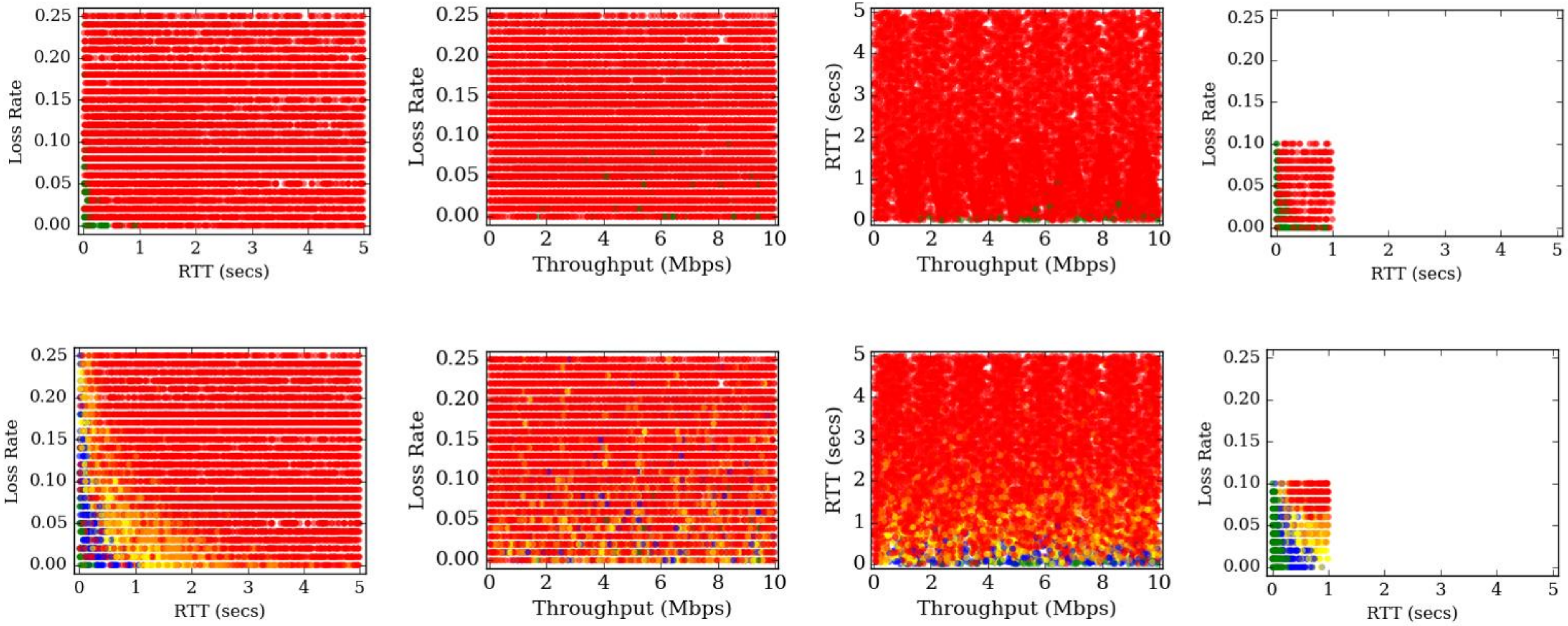
- **Instances Pool:**
  - RTT = 0 – 5000 ms
  - Loss Rate = 0 – 25 %
  - Throughput = 0 – 10 Mbps
- **Validation Set:**
  - RTT = 0 – 1000 ms
  - Loss Rate = 0 – 10 %
  - Throughput = 0 – 10 Mbps

Visual Representation of the datasets

Instances Pool **Validation Set**

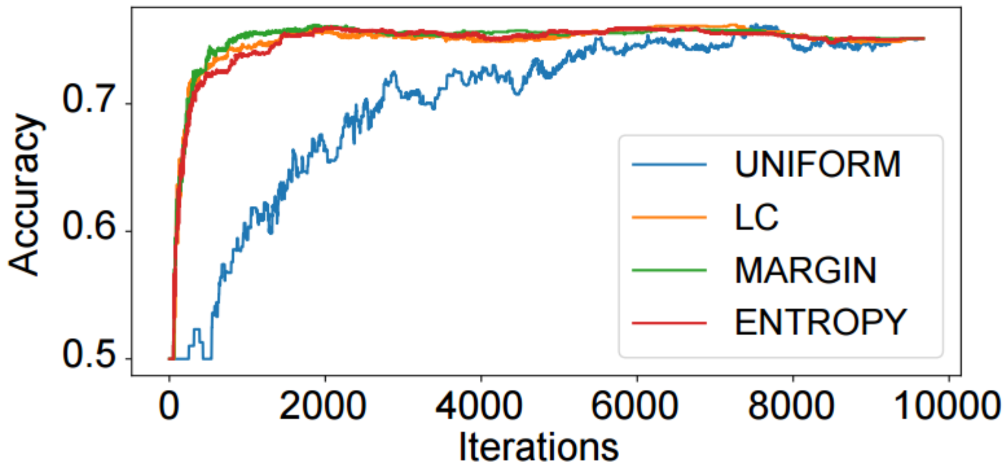
QoE_{binary}

QoE_{multi}



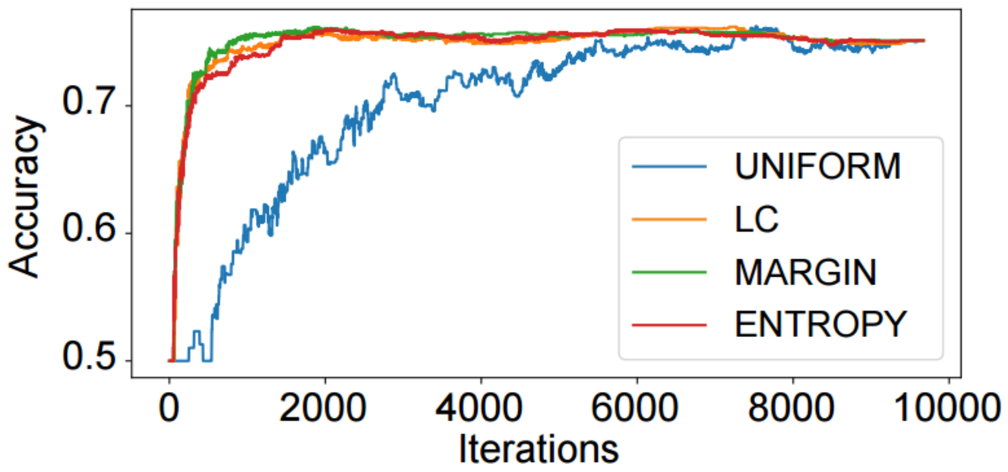
$RTT\ vs\ Loss\ Rate$ $Throughput\ vs\ Loss\ Rate$ $Throughput\ vs\ RTT$ $RTT\ vs\ Loss\ Rate$

Evaluation of our methodology – Binary Classification

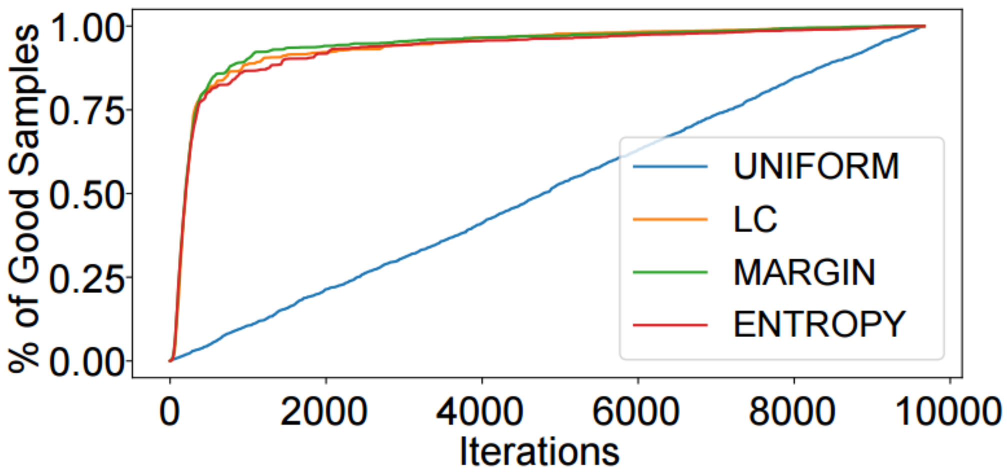


(a) Learner accuracy

Evaluation of our methodology – Binary Classification



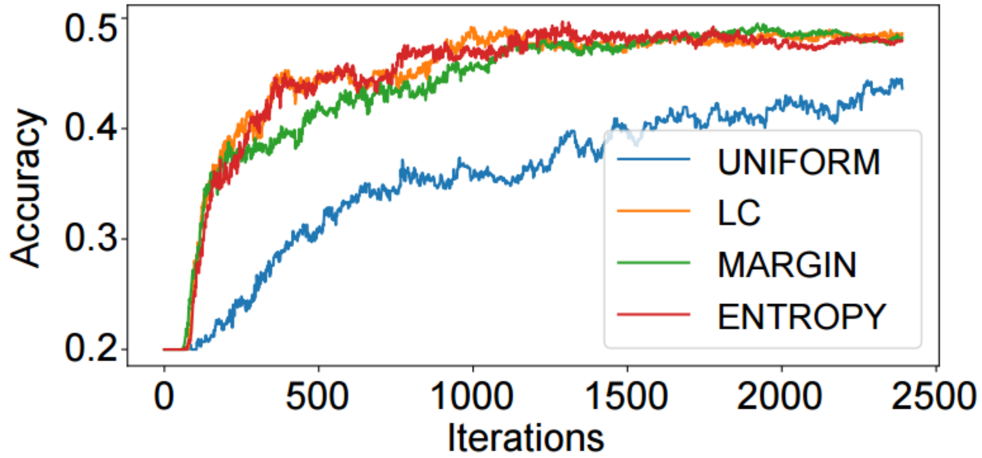
(a) Learner accuracy



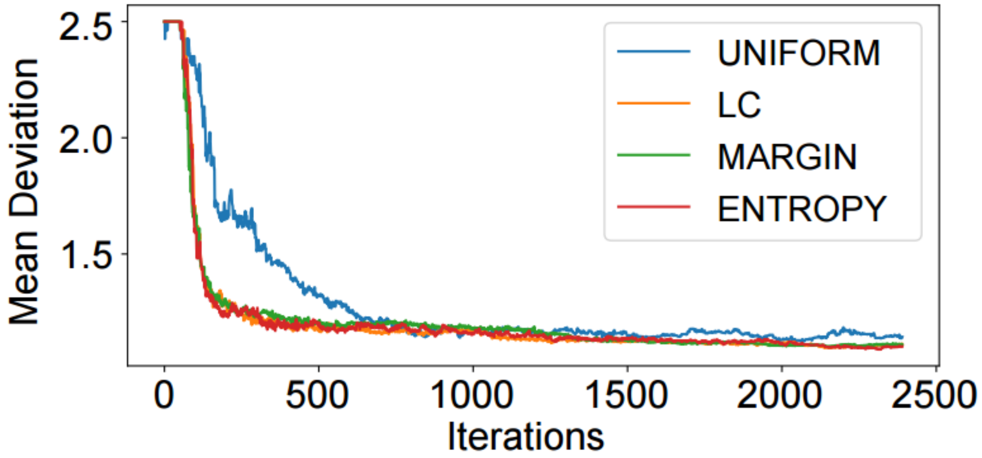
(b) Cumulative ratio of *Good* samples

INVENTEURS DU MONDE NUMÉRIQUE

Evaluation of our methodology – Multiclass Classification



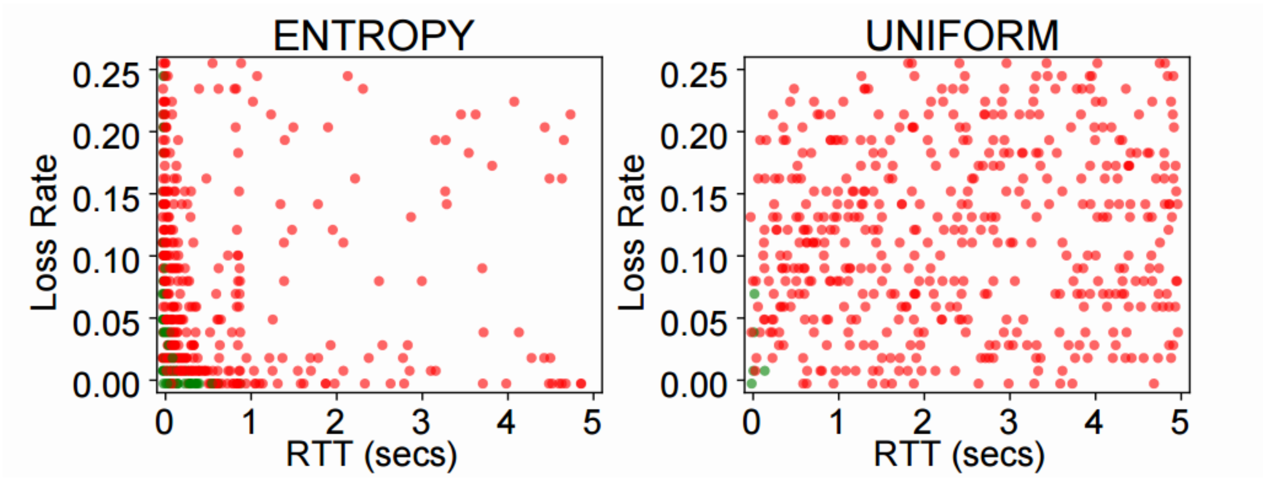
(a) Learner accuracy



(b) Mean Deviation

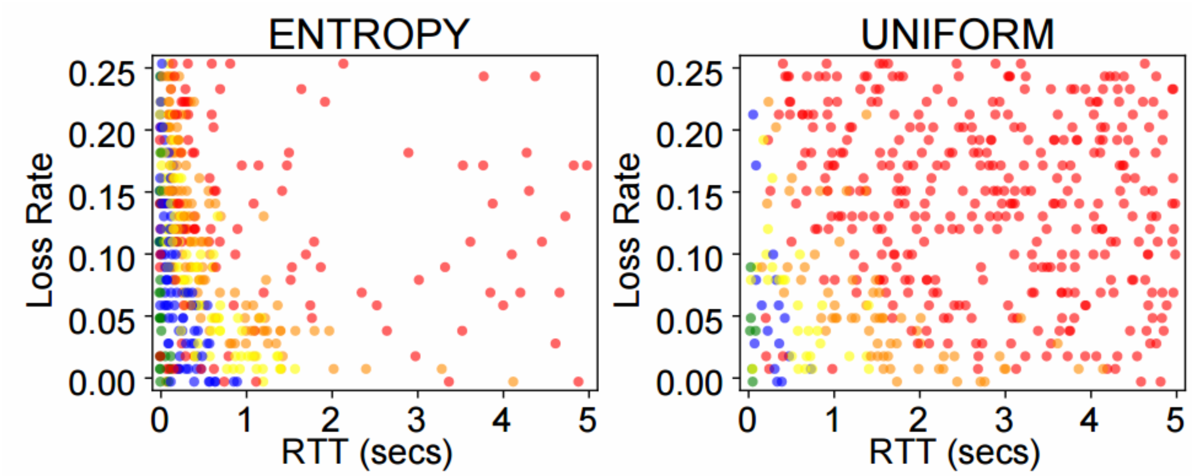
The Training sets at 5% of Pool size

QoE_{binary}



(a) Binary classification

QoE_{multi}



(b) Multiclass classification

Conclusions and future work

- Active learning provides a promising opportunity to speed up the process of building ML QoE models using controlled experimentation as shown in case of YouTube.
- Extend the work on more applications (e.g. Skype, Web etc.) and with more input features such as jitter, TCP re-ordering, etc.

Thank you!