Byzantine-tolerant federated Gaussian process regression for streaming data

Xu Zhang, Zhenyuan Yuan, Minghui Zhu

School of Electrical Engineering and Computer Science
Pennsylvania State University

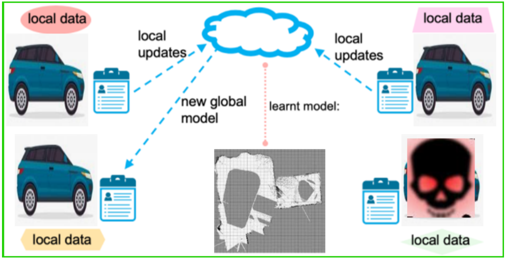
November 28, 2022

2022 Conference on Neural Information Processing Systems
New Orleans, LA

Problem formulation

Network model:

- Cloud can communicate with agents
- Agents cannot communicate with each other
- Byzantine agents send arbitrary model parameters to the cloud



Observation model:

$$y^{[i]}(t) = \quad (z^{[i]}(t)) + e^{[i]}(t)$$

- Training data $(z^{[i]}(t); y^{[i]}(t))$ arrive sequentially

Objective: Design a Byzantine-tolerant algorithm which

- Correctly learns the function
- Does not require to share local streaming data $(z^{[i]}(t); y^{[i]}(t))$

Byzantine-tolerant federated GPR

Contribution

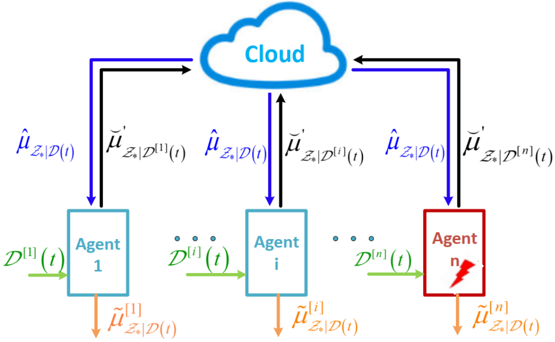
Design a Byzantine-tolerant federated Gaussian process regression (GPR) algorithm which

- Can guarantee the correct predictions and tolerate less than one quarter Byzantine agents
- Can deal with streaming data and perform on-line learning

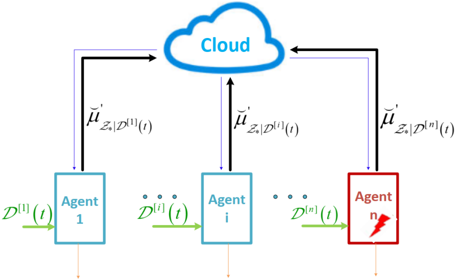
Agent-based local GPR

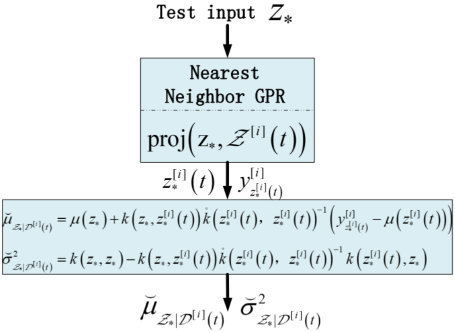
Cloud-based aggregated GPR

Agent-based fused GPR



Agent-based local GPR

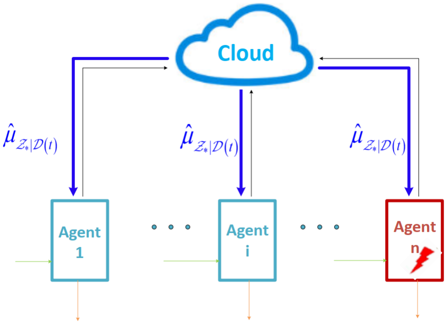


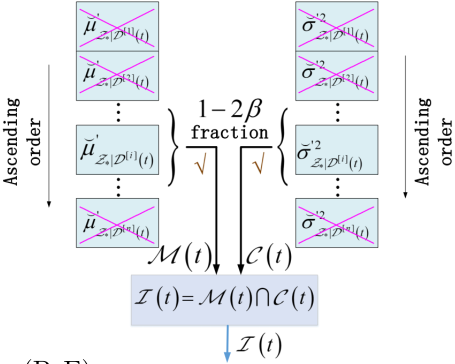


$\breve{\mu}'_{z|D^{[i]}(t)}$ $\breve{\mu}_{z|D^{[i]}(t)}$ Benign agent;
? Byzantine agent

$\breve{\sigma}'^2_{z|D^{[i]}(t)}$ $\breve{\sigma}^2_{z|D^{[i]}(t)}$ Benign agent;
? Byzantine agent

Cloud-based aggregated GPR



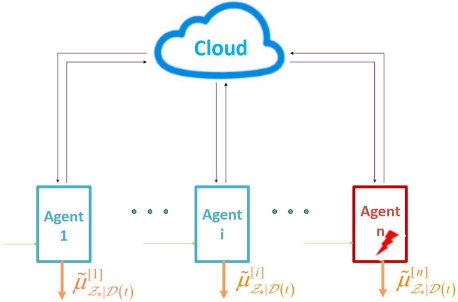


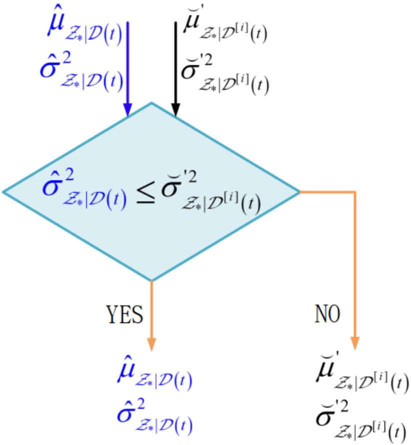
Byzantine-tolerant product of experts (PoE):

$$\hat{\mu}_{z|\mathcal{D}(t)} = \frac{\hat{\sigma}^2_{z|\mathcal{D}(t)}}{|\mathcal{I}(t)|}\sum_{i\in\mathcal{I}(t)} \breve{\mu}'_{z|\mathcal{D}^{[i]}(t)}\,\breve{\sigma}'^{-2}_{z|\mathcal{D}^{[i]}(t)},$$

$$\hat{\sigma}^2_{z|\mathcal{D}(t)} = \frac{|\mathcal{I}(t)|}{\sum_{i\in\mathcal{I}(t)} \breve{\sigma}'^{-2}_{z|\mathcal{D}^{[i]}(t)}}.$$

Agent-based fused GPR





Output: $\tilde{}^{[i]}_{z\ jD(t)}, (\tilde{}^{[i]}_{z\ jD(t)})^2$

Robustness of cloud-based aggregated GPR

Assumption

Less than one quarter of the agents are Byzantine.

Dispersion: $d^{[i]}(t) \triangleq \sup_{z \, 2 \, \mathcal{Z}} D(z; Z^{[i]}(t))$

Theorem (Cloud-based aggregated GPR: Mean)

For any $z \;\; 2 \, \mathcal{Z}$ *and* $0 < \quad < 1$*, with probability at least* $1 \quad$ *, it holds that*

$$\hat{\ }_{z \; jD(t)} \qquad (z \;\;)$$

$$(1 \quad \frac{(d^{\max}(t))}{{}^{2}_{f} + (\ {}^{\max}_{e})^{2}}) k \;\; k_{1} + \frac{{}^{2}_{f}\ \grave{}\ \ d^{\max}(t)}{{}^{2}_{f} + (\ {}^{\min}_{e})^{2}} + \mathrm{p}\overline{2 \;\; {}^{2}(\ln 2 \quad \ln \quad)} + \Delta(d^{\max}(t))$$ *where*

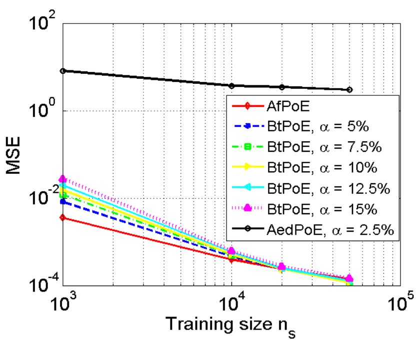
$$\Delta(s) \triangleq \frac{2 \;\; (\ \mathrm{p}\overline{2 \;\; {}^{2}(\ln(2n) \quad \ln \quad)} + \frac{{}^{2}_{f} k \;\; k_{1}}{{}^{2}_{f} + (\ {}^{\min}_{e})^{2}})}{1 \quad 4} \; \frac{{}^{4}_{f} + {}^{2}_{f}(\ {}^{\max}_{e})^{2} \qquad (s)^{2}}{{}^{2}_{f}(\ {}^{\min}_{e})^{2}}.$$

Theorem (Cloud-based aggregated GPR: Variance)

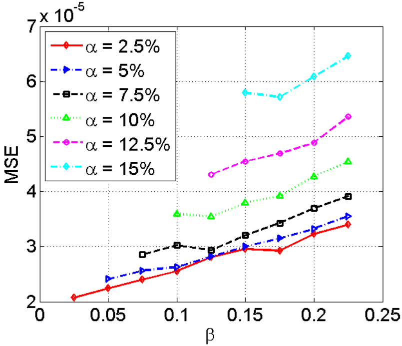
For any $z \;\; 2 \, \mathcal{Z}$*, it holds that* $\frac{{}^{2}_{f}(\ {}^{\min}_{e})^{2}}{{}^{2}_{f} + (\ {}^{\max}_{e})^{2}} \qquad \hat{\ }^{2}_{z \; jD(t)} \qquad {}^{2}_{f} \qquad \frac{(d^{\max}(t))^{2}}{{}^{2}_{f} + (\ {}^{\max}_{e})^{2}}.$

Experiments (Synthetic dataset)

Experiment 1: Prediction performance in terms of consistency and different ,



(a) Consistency evaluation



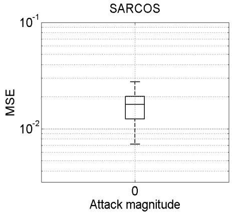
(b) Prediction performance on di erent β

Experiment 2: Prediction performance on different functions

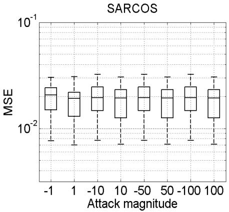
Algorithm	AfPoE	BtPoE	AedPoE
MSE (10 ^{3})	0:0049 0:007	0:0236 0:172	26:5339 0:019

Experiments (Real-world datasets)

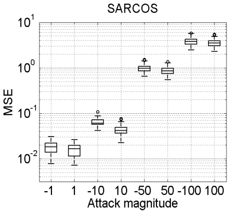
Experiment 3: Performance on different attack magnitudes



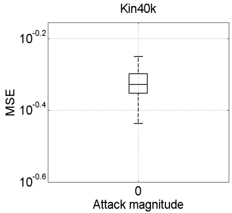
(c) Attack-free standard PoE



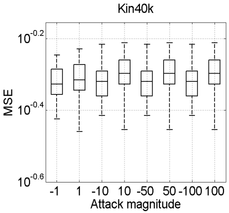
(d) Byzantine-tolerant PoE



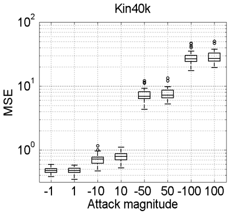
(e) Attacked standard PoE



(f) Attack-free standard PoE



(g) Byzantine-tolerant PoE



(h) Attacked standard PoE