ACM Multimedia 2022
Lisbon, Portugal | 10-14 October

Confederated Learning: Going Beyond Centralization

Zitai Wang[1,2], Qianqian Xu[3*], Ke Ma[4],

Xiaochun Cao[5,1], Qingming Huang[4,3,6,7*]

[1]SKLOIS, Institute of Information Engineering, CAS
[2]SCS, University of Chinese Academy of Sciences
[3]IIP, Institute of Computing Technology
[4]SCST, University of Chinese Academy of Sciences
[5]SCST, Sun Yat-sen University
[6]Key Lab. of BDKM, Chinese Academy of Sciences
[7]Peng Cheng Laboratory

Outline

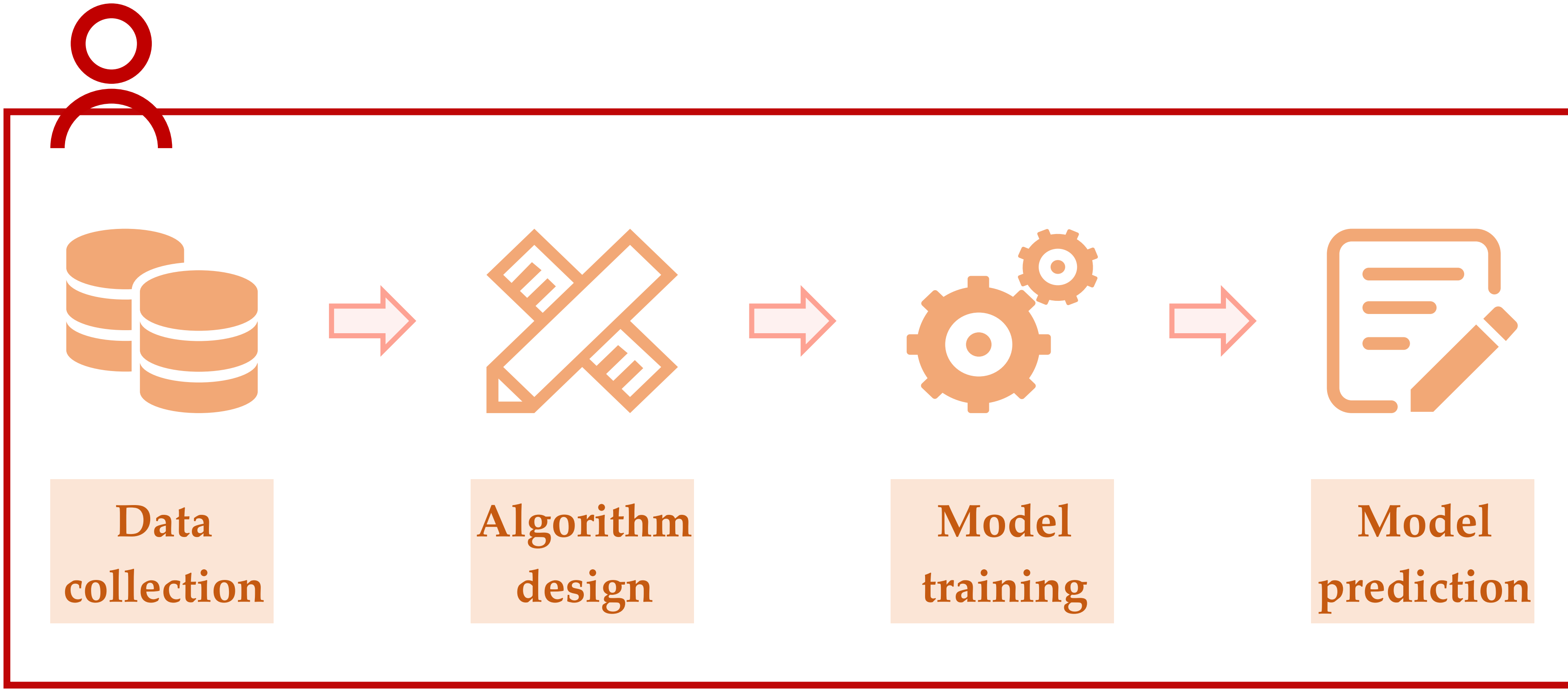
1 Background

2 Framework

3 Experiment

Background

- **Learning paradigm of traditional machine learning**



A single entity could control the whole learning process

Background

- **How to formulate traditional machine learning?**
  - ✓ **Training data and test data** $\mathcal{S} = \{\mathcal{S}^{\mathrm{tr}}, \mathcal{S}^{\mathrm{te}}\}$
  - ✓ **The evaluation metric defined on the test set** $\mathcal{M}$
  - ✓ **The hypothesis set** $\mathcal{H}$
  - ✓ **The specific learning algorithm** $\mathcal{A}$
  - ✓ **Then, we can formulate traditional machine learning as follows:**

$$\min_h \mathcal{M}(h; \mathcal{S}^{te})$$

where

$$h := \mathcal{A}(\mathcal{S}^{tr}, \mathcal{H})$$

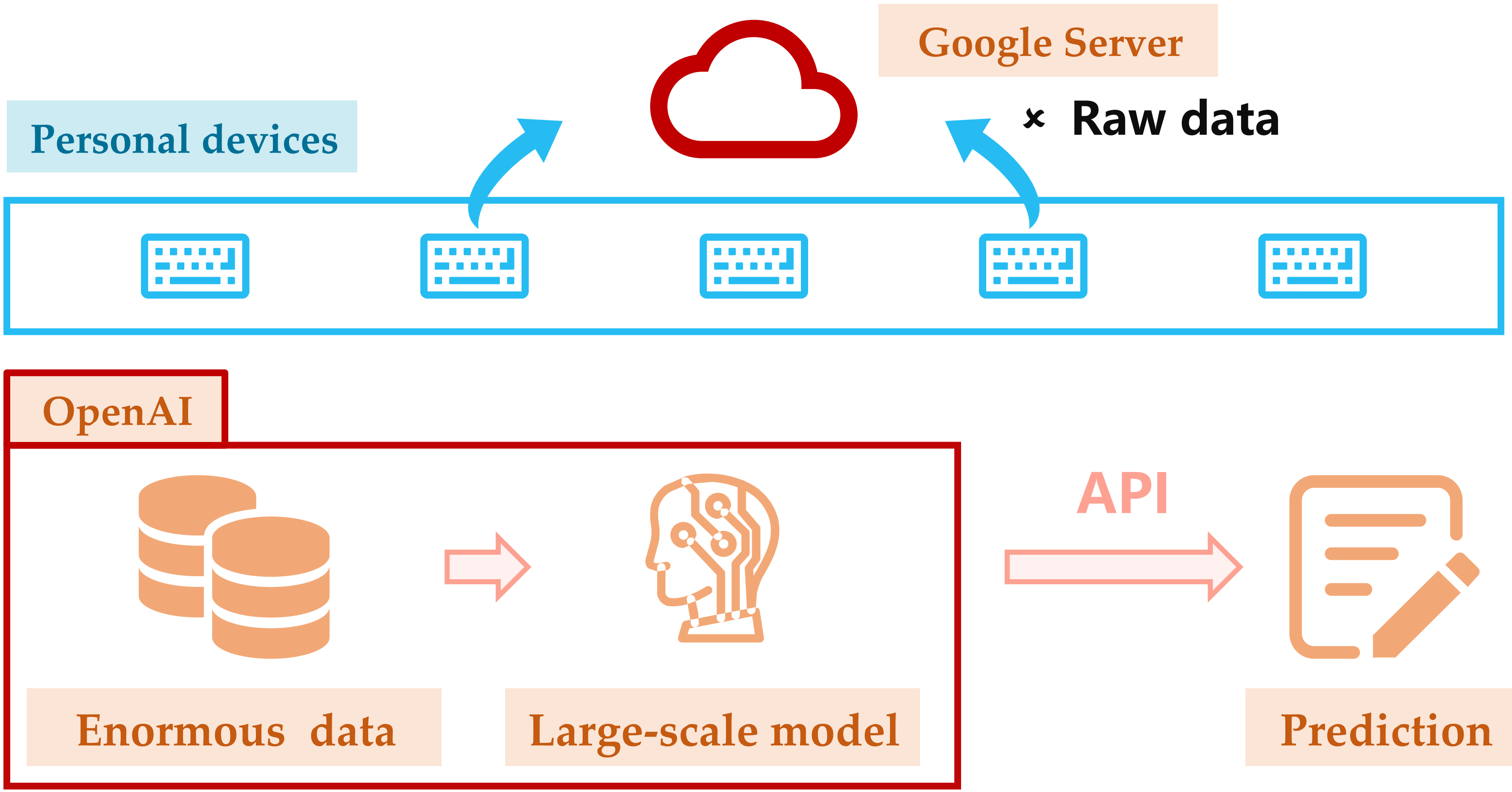
Background

- **An intuitive example: supervised learning**
  - ✓ **Training data** $\mathcal{S}^{\text{tr}} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$**, where** $\boldsymbol{x}_i$ **is the input drawn from the feature space** $\mathcal{X}$**, and** y_i **denotes the associated label drawn from the label space** $\mathcal{Y}$
  - ✓ **The hypothesis set** $\mathcal{H} = \{h : \mathcal{X} \to \mathcal{Y}\}$
  - ✓ **The specific learning algorithm: GSD, Adam…**
  - ✓ **The evaluation metric: Accuracy, AUC, mAP, NDCG…**
  - ✓ **Then, Empirical Risk Minimization:**

$$\min_h \sum_{i=1}^{m} \ell(h(\boldsymbol{x}_i), y_i) + \Omega(h)$$

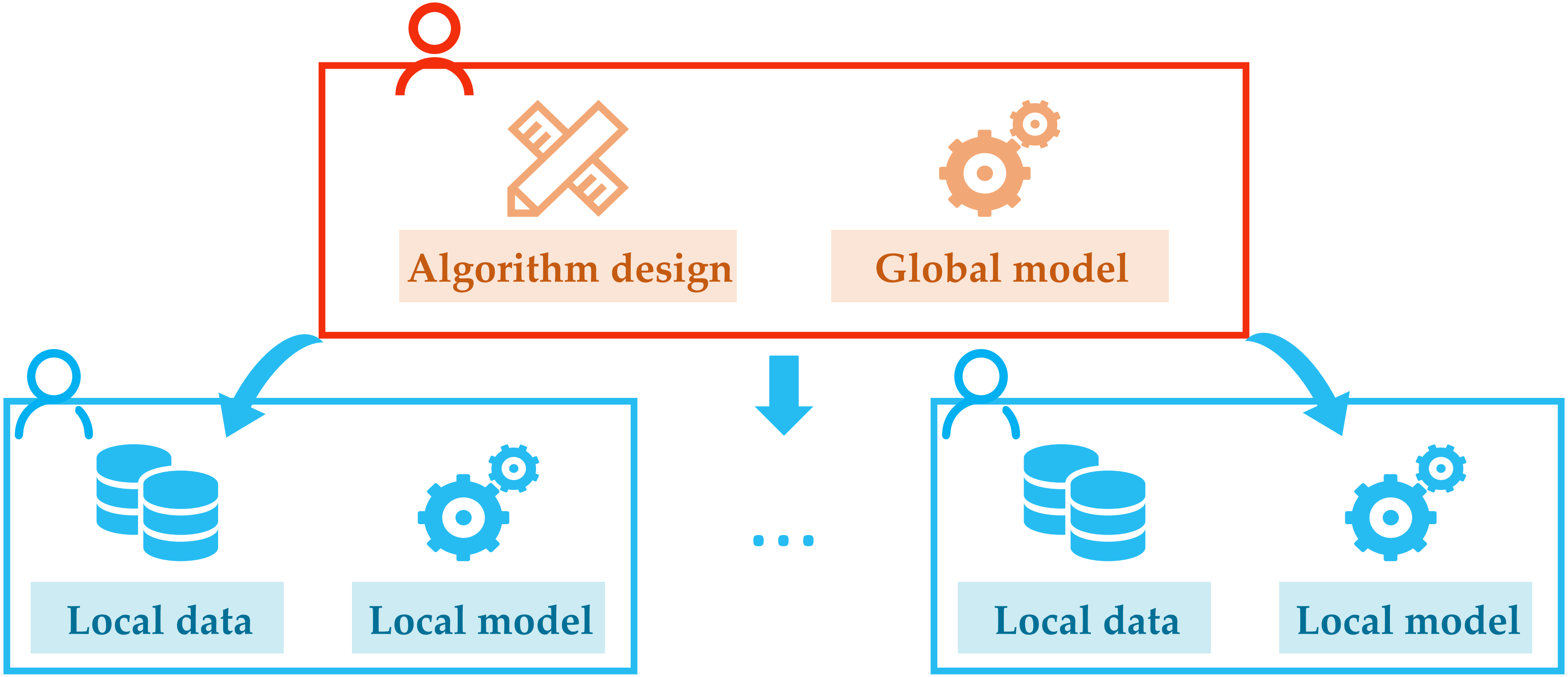
Background

- Cooperation among entities become crucial due to cost, privacy and security concerns



Background

□ Learning paradigm of federated learning



The center entity controls the whole learning process

Background

□ How to describe federated learning?

- ✓ **Training data and test data** $\mathcal{S} = \cup\{\mathcal{S}_i^{\mathrm{tr}}, \mathcal{S}_i^{\mathrm{te}}\}_{i=1}^{N_{eg}}$
- ✓ **The hypothesis set** $\mathcal{H}$
- ✓ **The specific learning algorithm** $\mathcal{A} = \{\mathcal{A}^l, \mathcal{A}^g\}$
- ✓ **The evaluation metric defined on the test set** $\mathcal{M}$
- ✓ **Besides, two types of entities are necessary:**

$$\mathcal{E} = \{e_i^{\mathrm{eg}}\}_{i=1}^{N_{\mathrm{eg}}} \cup \{e^{\mathrm{ct}}\}$$

Background

- **Then federated learning consists of four steps:**
  - ✓ **The center entity e^{ct} design the algorithm $\mathcal{A}$ and initializes the global model from $\mathcal{H}$**
  - ✓ **The center entity transmits the current model to the edge entities**
  - ✓ **Each edge entity e_i^{eg} updates the received model based on the local data $\mathcal{S}_i^{\mathrm{tr}}$ via the local update algorithm $\mathcal{A}^l$**
  - ✓ **The center entity collects the local models and**

It is necessary to consider different entities

Background

- **However..**
  - It only provides two types of entities: center and edge. Such a rigid role setting fails to cover many cooperation scenarios
  - The learning process depends on the credibility of the center entity, while establishing a trustworthy center entity is generally costly

A more generalized learning paradigm for cooperation?

Outline

1 Background

2 Framework

3 Experiment

Framework

- Inspired by the concept of permission of database, we additionally define
  - ✓ $\mathcal{R} = \{r_i\}_{i=1}^{N_r}$, the set of roles (i.e., different types of entities)
  - ✓ $\mathcal{P} = \{p_n\}_{n=1}^{N_p}$, the set of permissions (e.g. create, read, delete, update…)
  - ✓ Then, $p_n(r_i, \mathcal{S}_j)$ means the role r_i has the permission p_n on $\mathcal{S}_j$
  - ✓ $\mathcal{P}_\mathcal{R} \subset \mathcal{P} \times \mathcal{R}$ denotes all the roles' permissions

Framework

ACM Multimedia 2022
Lisbon, Portugal | 10-14 October

- Then, Cooperative learning problem can be described by:

$$\{\mathcal{R}, \mathcal{P}, \mathcal{P}_{\mathcal{R}}, \mathcal{S}, \mathcal{H}, \mathcal{A}, \mathcal{M}\}$$

- For traditional machine learning:
  - ✓ There only exists a single role $\mathcal{R} = \{r\}$
  - ✓ r naturally has permissions to all the factors
  - ✓ We can formulate traditional machine learning as

$$\{r, \mathcal{P}, \mathcal{P}_{r}, \mathcal{S}, \mathcal{H}, \mathcal{A}, \mathcal{M}\}$$

Framework

- **For federated learning:**
  - ✓ **Let r^{eg} denote the edge role and be the $\{r_i^{\mathrm{eg}}\}_{i=1}^{N_{\mathrm{eg}}}$ corresponding entities**
  - ✓ **Then,**

$$\left\{\{r^{\mathrm{eg}}, r^{\mathrm{ct}}\}, \mathcal{P}, \mathcal{P}_{\mathcal{R}}, \{\mathcal{S}_i\}_{i=1}^{N_{\mathrm{eg}}}, \mathcal{H}, \{\mathcal{A}^g, \mathcal{A}^l\}, \mathcal{M}\right\}$$

where

$$P(r_i^{\mathrm{eg}}) = \{p(r^{\mathrm{eg}}, \theta_{\varnothing,\mathcal{H},\mathcal{A}^g}), p(r^{\mathrm{eg}}, \mathcal{S}), p(r^{\mathrm{eg}}, \mathcal{A}^l), p(r^{\mathrm{eg}}, \mathcal{M})\}$$

$$P(r^{\mathrm{ct}}) = \{p(r^{\mathrm{ct}}, \theta_{\mathcal{S}^{\mathrm{eg}},\mathcal{H},\mathcal{A}^l}), p(r^{\mathrm{ct}}, \mathcal{H}), p(r^{\mathrm{ct}}, \mathcal{A}^g), p(r^{\mathrm{ct}}, \mathcal{A}^l)\}$$

$$p(r^{\mathrm{eg}}, \mathcal{S}) = \{p(r_i^{\mathrm{eg}}, \mathcal{S}_i)\}_{i=1}^{N_{\mathrm{eg}}}, p(r^{\mathrm{ct}}, \theta_{\mathcal{S}^{\mathrm{eg}},\mathcal{H},\mathcal{A}^l}) = \{p(r^{\mathrm{ct}}, \theta_{\mathcal{S}_i^{\mathrm{eg}},\mathcal{H},\mathcal{A}^l})\}_{i=1}^{N_{\mathrm{eg}}}$$

Framework

- **Assume that**
  - ✓ **A cloud platform (i.e., r_1) provides its computing power with a non-customizable algorithm $\mathcal{A}_1$ and a predetermined hypothesis set $\mathcal{H}_1$**
  - ✓ **How should we (i.e., r_2) collect training samples $\mathcal{S}_2$ according to the current performance on metric set $\mathcal{M}_2$?**
  - ✓ **How to formulate this problem?**

Framework

- **According to the proposed framework, the problem could be denoted as**

$$\{\{r_1, r_2\}, \mathcal{P}, \mathcal{P}_{\mathcal{R}}, \mathcal{S}_2, \mathcal{A}_1, \mathcal{H}_1, \mathcal{M}_2\}$$

where

$$\mathcal{P}_{\mathcal{R}} = \{P(r_1), P(r_2)\}$$

$$P(r_1) = \{p(r_1, \mathcal{H}_1), p(r_1, \mathcal{A}_1)\}$$

$$P(r_2) = \{p(r_1, \theta_{2,1,1}), p(r_2, \mathcal{S}_2), p(r_2, \mathcal{M}_2)\}$$

Framework

- For the problem we discussed before (GPT-3):
  - ✓ r^a the role providing API
  - ✓ r^T the target role, aims to improve the performance on the target set $\mathcal{S}^T$
  - ✓ Then,

$$\{\{r^a, r^T\}, \mathcal{P}, \mathcal{P}_{\mathcal{R}}, \{\mathcal{S}^a, \mathcal{S}^T\}, \{\mathcal{H}^a, \mathcal{H}^T\}, \{\mathcal{A}^a, \mathcal{A}^T\}, \mathcal{M}^T\}$$

where

$$P(r^a) = \{p(r^a, \mathcal{S}^a), p(r^a, \mathcal{H}^a), p(r^a, \mathcal{A}^a)\}$$

$$P(r^T) = \{p(r^T, \theta_{\mathcal{S}^a, \mathcal{H}^a, \mathcal{A}^a}), p(r^T, \mathcal{S}^T), p(r^T, \mathcal{H}^T), p(r^T, \mathcal{A}^T), p(r^T, \mathcal{M}^T)\}$$

$$p(r^T, \theta_{\mathcal{S}^a, \mathcal{H}^a, \mathcal{A}^a}) = \text{'predictions'}$$

Framework

- **We assume that the target dataset**

$$\{(\boldsymbol{x}_i, \overline{y}_i)\}_{i=1}^{m}$$

suffers from a distribution mismatch, where $\overline{y}_i$ denotes the observed labels.

- **How to eliminate the distribution mismatch with the help of r^a ?**
  - ✓ **Sample Reweighting**
  - ✓ **Label Ensemble**
  - ✓ **Consistent Regularization**

Framework

ACM Multimedia 2022
Lisbon, Portugal | 10-14 October

☐ Sample Reweighting

$$\theta^* = \arg\min_\theta \sum_{i=1}^{m} h_w \ell_i(\theta) + \Omega_\theta$$

$$h_w(g_i^*, \overline{y}_i) = \frac{\exp(-\alpha \cdot d(g_i^*, \overline{y}_i))}{\sum_{i=1}^{m} \exp(-\alpha \cdot d(g_i^*, \overline{y}_i))},$$

☐ Label Ensemble

$$\hat{y}_i \leftarrow \beta \overline{y}_i + (1-\beta) g_i^*$$

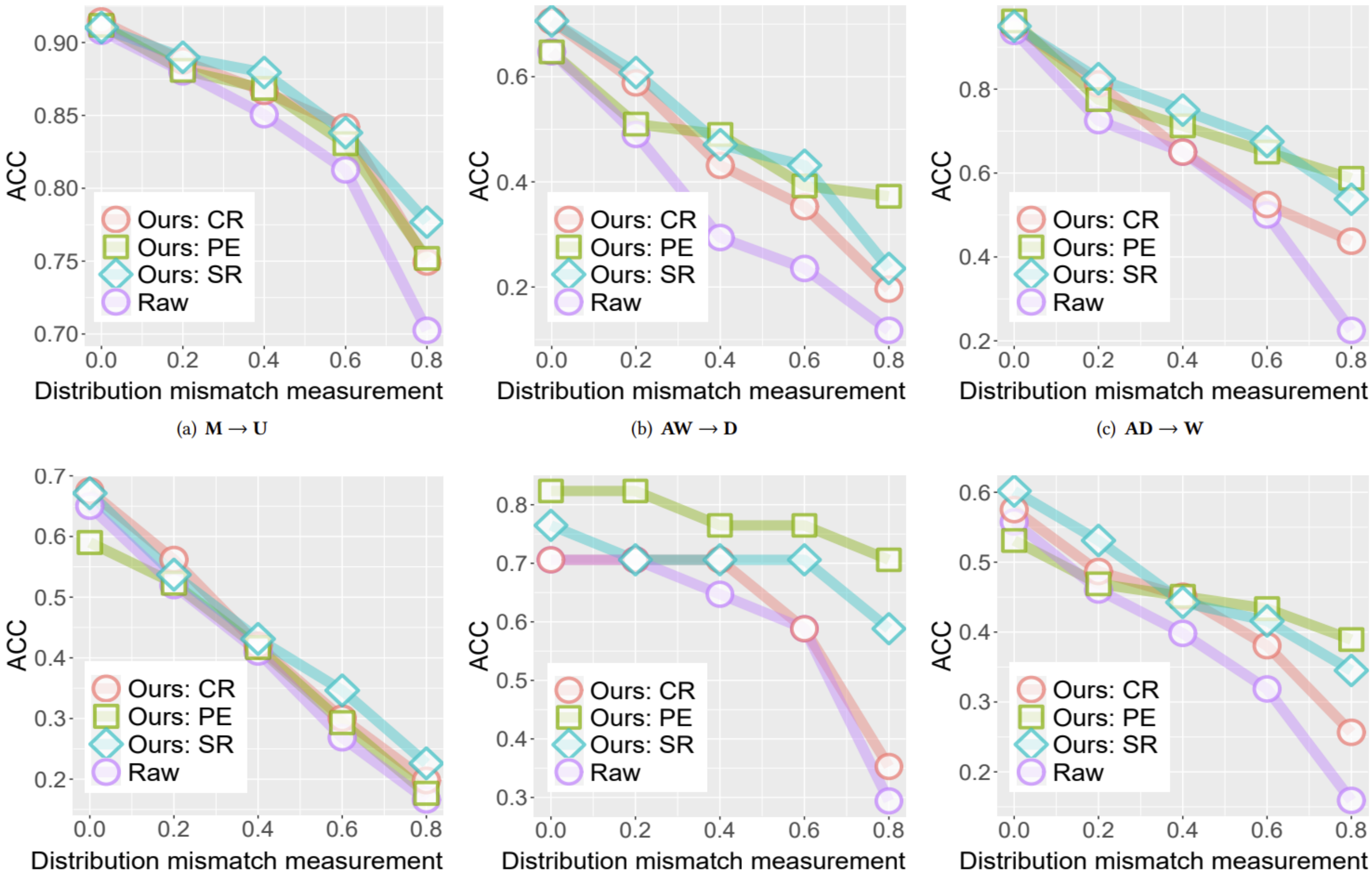
☐ Consistent Regularization

$$\Omega_D(g_i^*, f_i(\theta)) = \gamma \sum_{i=1}^{m} \| g_i^* - f_i(\theta) \|^2$$

Outline

1. Background
2. Framework
3. Experiment

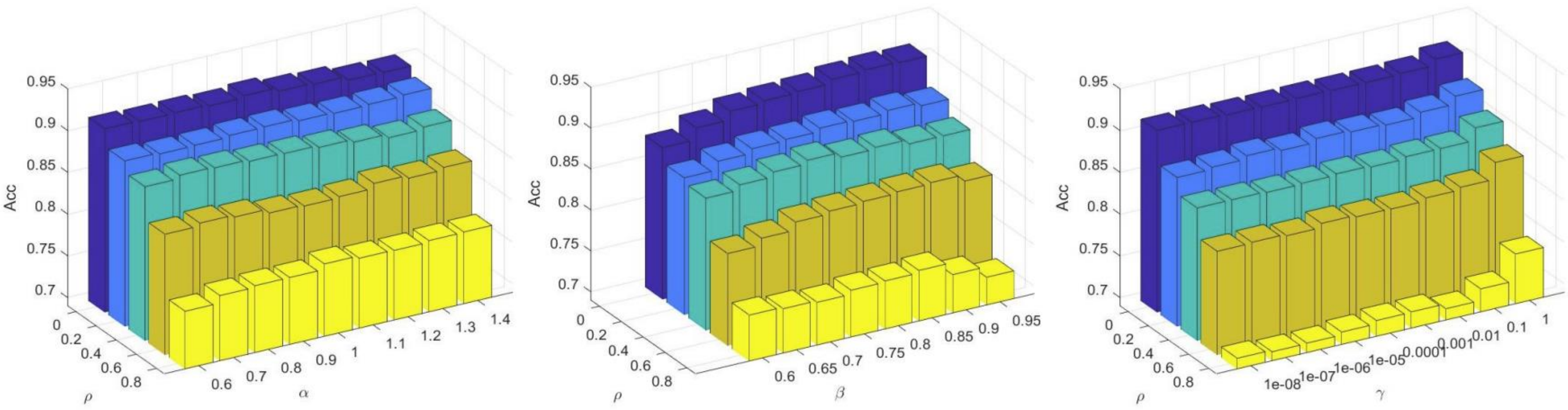
Experiment



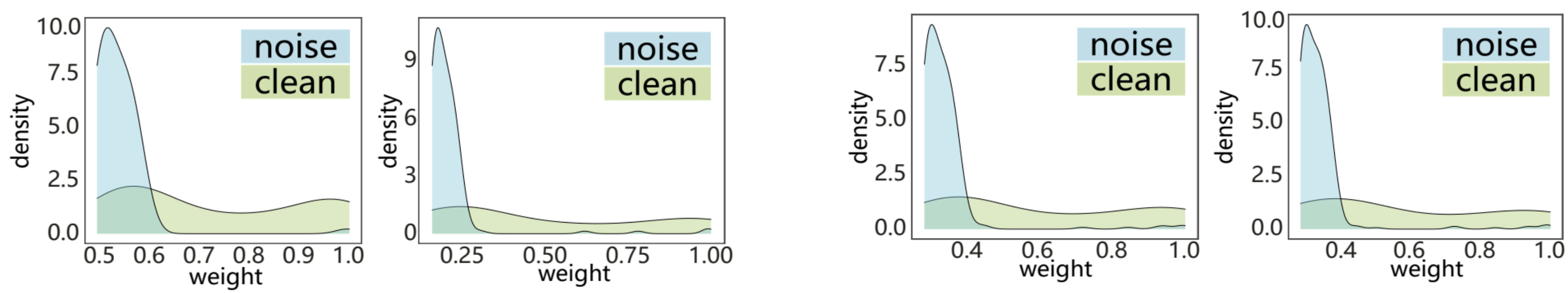
(a) $\mathbf{M} \rightarrow \mathbf{U}$ (b) $\mathbf{AW} \rightarrow \mathbf{D}$ (c) $\mathbf{AD} \rightarrow \mathbf{W}$

(d) $\mathbf{AD} \rightarrow \mathbf{W}$ (e) $\mathbf{ACW} \rightarrow \mathbf{D}$ (f) $\mathbf{ADW} \rightarrow \mathbf{C}$

Experiment



(a) SR (b) PE (c) CR



(a) $\rho = 0.2$ (b) $\rho = 0.4$ (c) $\rho = 0.6$ (d) $\rho = 0.8$