Towards Collaborative Multi-modal Federated Learning for Human Activity Recognition in Smart Workplace Environments

S. Mohammad Sheikholeslami[1], Pai Chet Ng[1], Huan Liu[2], Yanhao Yu[2], Konstantinos N. Plataniotis[1]

[1] Department of Electrical and Computer Engineering, University of Toronto
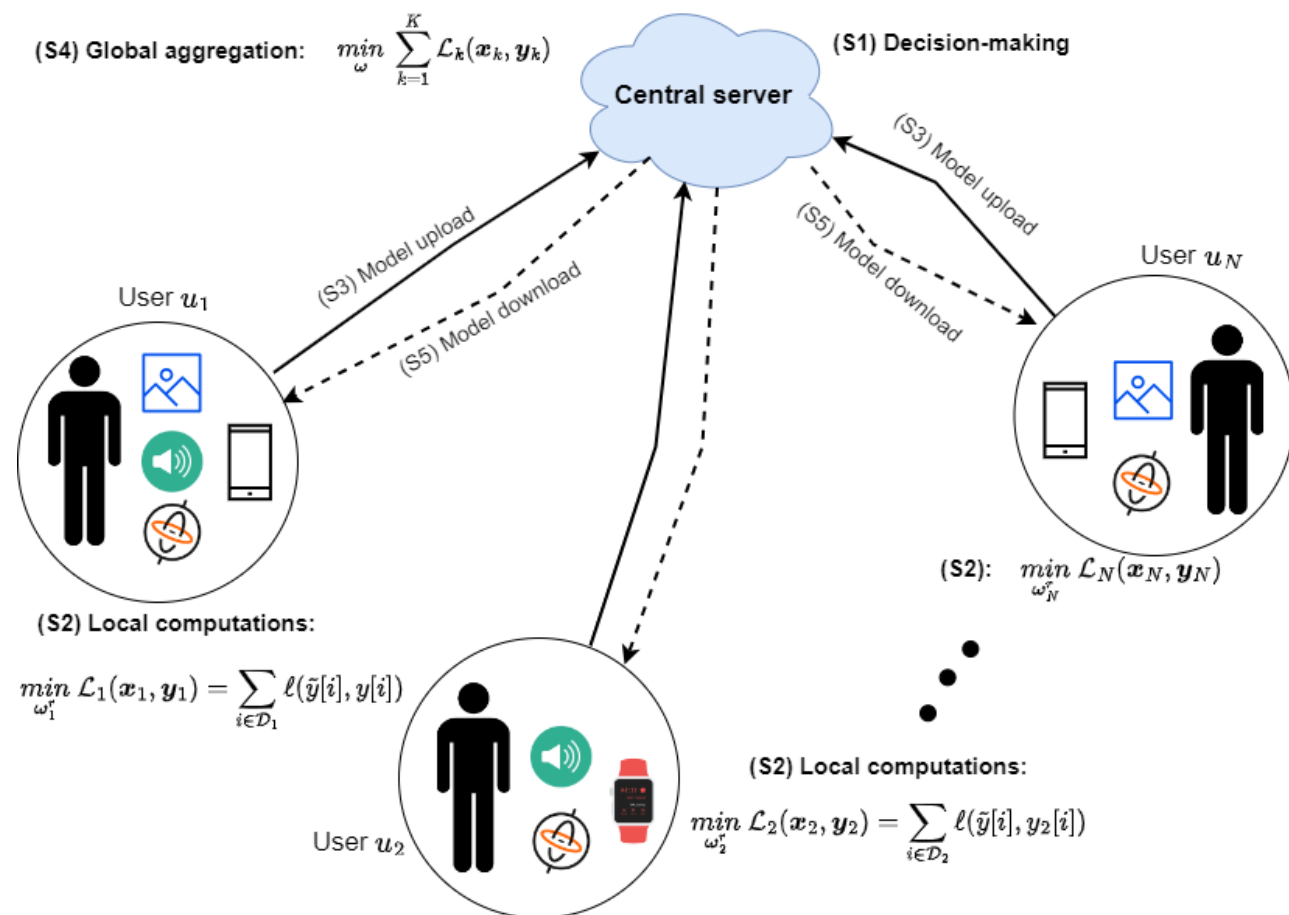
[2] Huawei Noah's Ark Lab, Canada

⋙Outline

❑ Introduction and Motivation

❑ Contributions

❑ System model and proposed Collaborative Multi-modal Federated Learning (CoMFL)

❑ Experimentations and results

➢ Data over networks comes
  ▪ in form of different modalities
  ▪ Is distributed over multiple devices

➢ **RQ1:** How can data distributed over multiple devices of a single user be combined to create a super model?

➢ **RQ2:** How can private data from different user can be combined to train a richer super model?

(S4) Global aggregation: $\min_{\omega} \sum_{k=1}^{K} \mathcal{L}_k(\boldsymbol{x}_k, \boldsymbol{y}_k)$

(S1) Decision-making

Central server

(S3) Model upload

(S5) Model download

User u_1

(S3) Model upload

(S5) Model download

User u_N

(S2): $\min_{\omega_N^\tau} \mathcal{L}_N(\boldsymbol{x}_N, \boldsymbol{y}_N)$

(S2) Local computations:

$$\min_{\omega_1^\tau} \mathcal{L}_1(\boldsymbol{x}_1, \boldsymbol{y}_1) = \sum_{i \in \mathcal{D}_1} \ell(\tilde{y}[i], y[i])$$

(S2) Local computations:

$$\min_{\omega_2^\tau} \mathcal{L}_2(\boldsymbol{x}_2, \boldsymbol{y}_2) = \sum_{i \in \mathcal{D}_2} \ell(\tilde{y}[i], y_2[i])$$

User u_2

Contributions

1) Local (Intra-zone) training: Collaboration of multiple devices with different sensing inputs belonging to the same user to train a local multimodal HAR model

2) Global (Inter-zone) training: the cross-user cooperation of different users enriches the global model without sharing their private multimodal data

3) A robust feature reconstruction network to adaptively reconstruct the missing data.

4) a comparative analysis between unimodal and multimodal approaches within both centralized and distributed settings with missing features
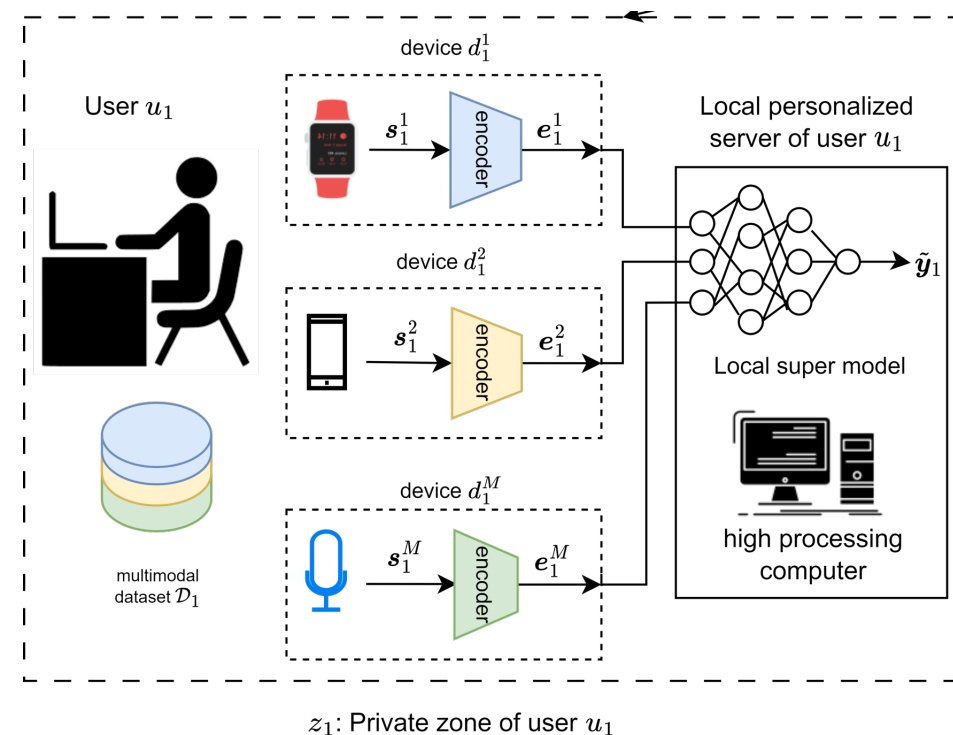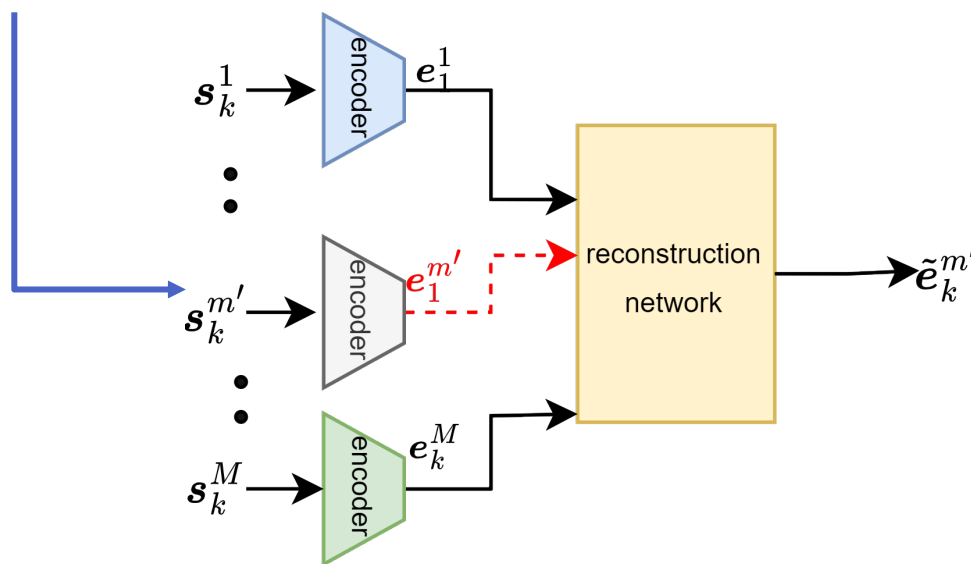
Multi-modal data fusion(RQ1): intermediate fusion

Private zone z_k: the set of devices belonging to user u_k

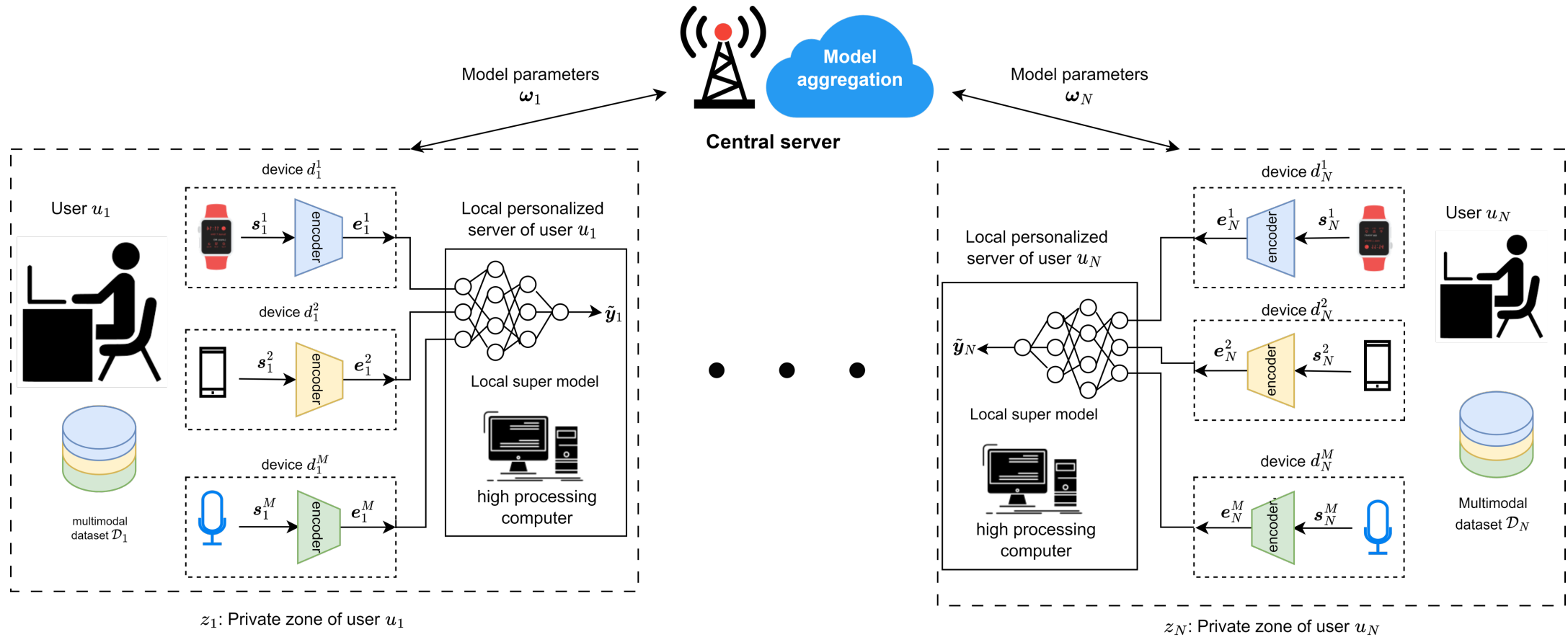Dataset: Multi-modal human activity recognition

Problem of intermediate fusion: missing modality in the inference

Proposed solution

z_1: Private zone of user u_1

Collaborative Multi-modal Federated Learning (RQ2)

Experimentation setup)

Multi-modal HAR dataset

➢ Each user possesses
  • Smartphone
  • Smartwatch
  • Smartspeaker
  • Computer

➢ Eight activity classes

➢ Total 28 users:
  ▪ 20 active users: participate in training
  ▪ 8 passive users: do not participate in training

**Table 1**: List of activities in the multimodal HAR dataset.

Simple activities	Complex activities
(a1) walking around (a2) talking on the phone (a3) reading book (a4) stretching (a5) writing (a7) browsing the monitor (a6) typing on the keyboard	(a8) walking and talking on the phone (a9) writing and talking on the phone

Experiment 1)

Test accuracy vs Communication rounds (epochs) for active users

- In general, central training outperforms federated learning

- Phone modality provides the best accuracy among unimodal HAR models in both distributed and central settings

- Multi-modal fusion enhances performance of the HAR models
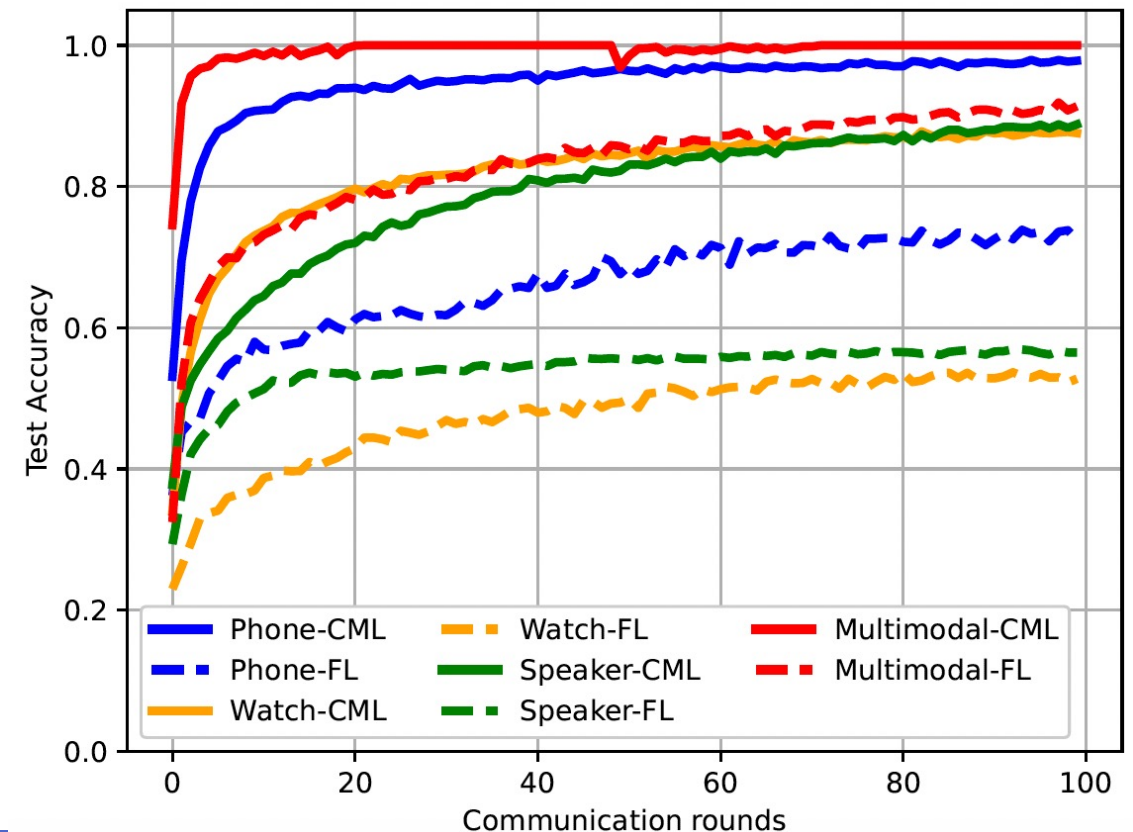
Experiment 2)

Performance on cross-domain data (passive users)

- The performance of the models trained over actives users reduces significantly over passive users

- Results illustrate weak generalization of local models

- The propose CoMFL enhances generalization

- CoMFL performs almost same as centralized case while preserving the privacy of users
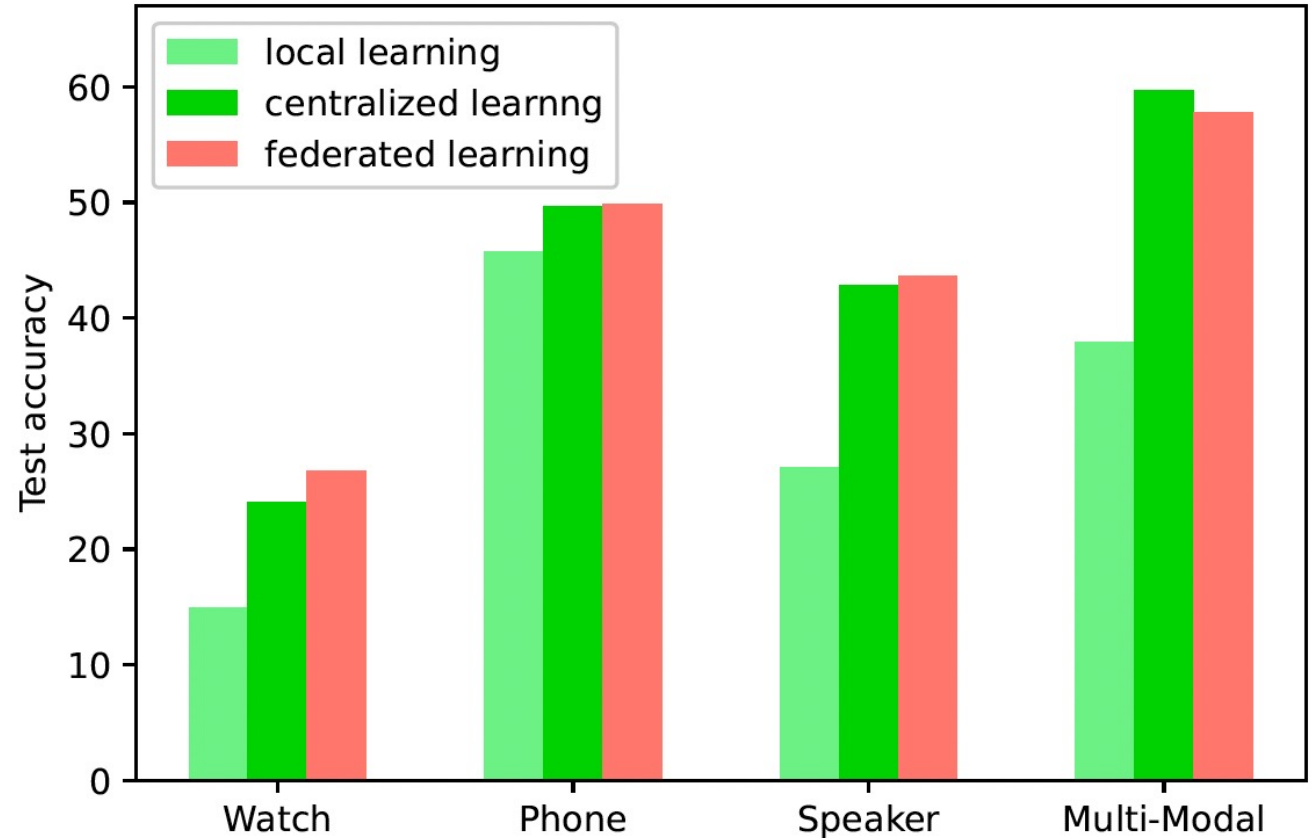
Experiment 3)

Effect of missing modality

- Phone modality has the greatest effect on the performance of the super model when missing

- The proposed reconstruction model recovers the missing modality and enhances performance of the model

Table 2: Performance of the multimodal HAR models with missing modality.

training strategy	missing modality						full modality
	watch		phone		speaker		
	baseline	proposed	baseline	proposed	baseline	proposed	
MM centralized learning model	83.21%	90.83%	38.11%	46.17%	94.54%	98.14%	99.29%
MM federated learning model	74.03%	78.44%	28.33%	54.26%	79.21%	85.79%	91.37%

Thank you for your attention

Questions?