Federated Multi-Task Attention for Cross-Individual Human Activity Recognition

Qiang Shen[1], Haotian Feng[1], Rui Song[2], Stefano Teso[3], Fausto Giunchiglia[1,3] and Hao Xu [1,2]

1 College of Computer Science and Technology, Jilin University
2 School of Artificial Intelligence, Jilin University
3 University of Trento

Motivation

Cross-individual distribution discrepancy in HAR

Statistical Perspective:

People are characterized by different habits, lifestyles and behavior patterns, which means that **the same activity may be performed very differently by different individuals, inducing a substantial cross-individual discrepancy** in the conditional distribution of activities given sensor observations.

System Perspective:

A major consequence of this fact is that, from the perspective of HAR applications, it is challenging to leverage statistical models learned on known users, for which annotated data is available, for predicting the activity of new users with their own activity characteristics.

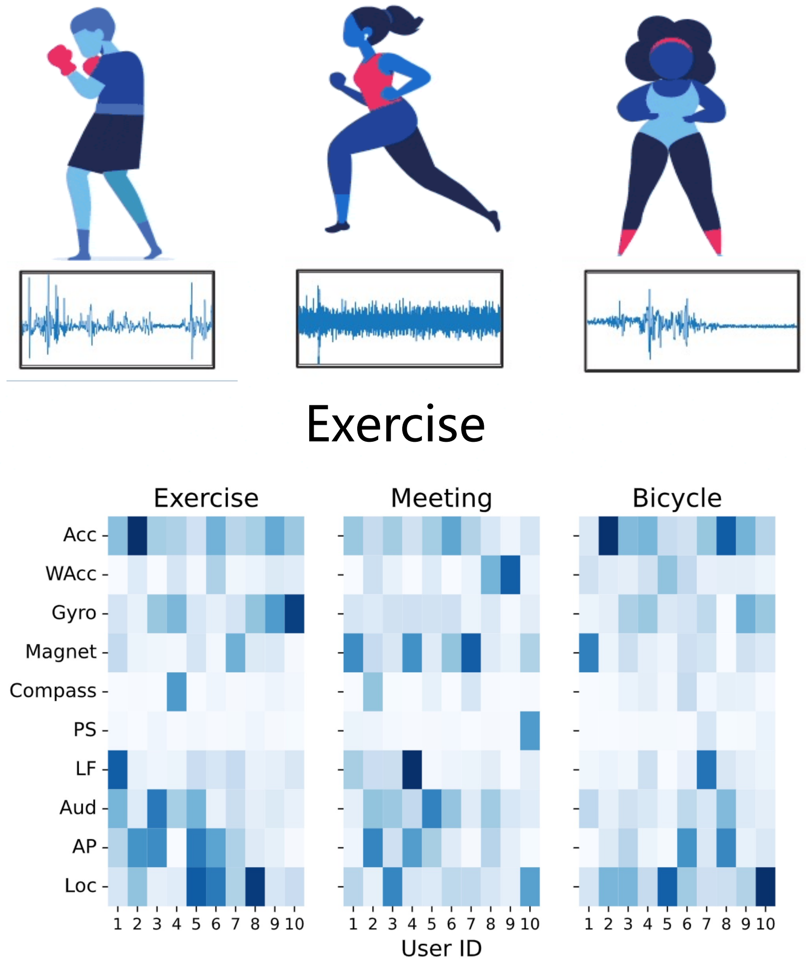

Figure 1: Importance of different features for 3 activities from 10 different individuals in ExtraSensory dataset. Saturation indicates higher relevance. The images indicate that the features important for recognizing any given activity strongly depend on the target user.

Method

Overview Architecture

☐ Federated multi-task framework

Extracts and fuses individual-agnostic and individual-specific multimodal features in a federated multi-task learning manner.

☐ Multi-task attention mechanism

Works as a mask for learning individual-specific features from the shared model while allowing for features to be shared among different individuals.

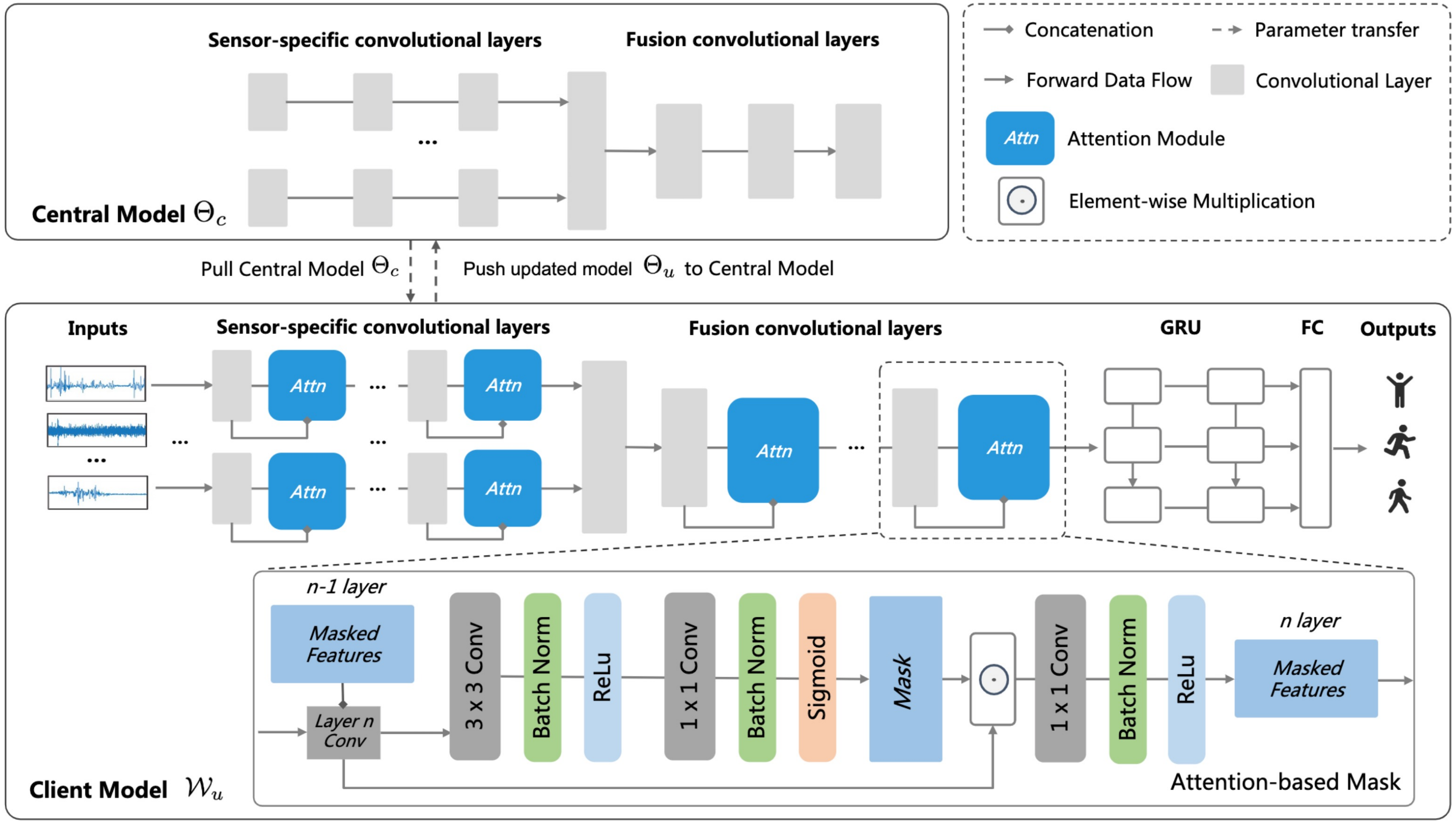

Figure 2: Architecture of FedMAT. Structures of the central model and one of the client models are visualized.

Method

Objective

The proposed architecture consists of a central model, with parameters Θ_c, and m decentralized models $\mathcal{W}_u, u \in \{1, 2, \cdots, m\}$ that learn individual-specific features. The overall goal is to acquire a HAR model that generalizes (i) across observed individuals, represented by $\mathcal{U}$, and (ii) to new individuals outside of $\mathcal{U}$.

$$\min_{\Theta_c, \mathcal{W}_u} \sum_{u=1}^{m} \sum_{i=1}^{n_u} l_u(f_u(x_u^i; \Theta_c, \mathcal{W}_u), y_u^i). \tag{1}$$

$$\Theta_c = \Theta_c + \lambda(\hat{\Theta} - \Theta_c), \tag{2}$$

$$\hat{\Theta} = \frac{1}{m} \sum_{u=1}^{m} \Theta_u. \tag{3}$$

Federated Model Update

Algorithm 1 FedMAT.

Input: m individual-specific data sets $\{\mathcal{D}_u\}$, one per client.
Output: central model Θ_c, individual-specific models $\{\mathcal{W}_u\}$.

```
Server: Initialize central model Θ_c ← Θ_0
for round = 1, 2, ... do
    for each u ∈ {1, 2, ..., m} in parallel do
        Client u: Get central model Θ_c from the server.
        Client u: Train for n epochs using central model
        Θ_c together with local model W_u, and get locally
        updated parameters Θ_u and W_u.
        Client u: Push updated parameters Θ_u to server.
    end for
    Server: Update Θ_c according to Eq. 2
end for
return Θ_c and {W_1, ..., W_m}
```

Method

Attention-based Mask

We apply the attention-based mask to the feature representation layers, aiming at extracting individual-specific information.
Specifically, we refer the shared features in the *l-th* layer of the shared network as e^l, and the learned attention mask in this layer for individual u as e_u^l . The task-specific features $\hat{e}_u^l$ in this layer, are then computed by element-wise multiplication of the attention masks with the shared features:

$$\hat{e}_u^l = Mask_u^l \odot p^j. \qquad (4)$$

For the attention mask in layer j, the input the concatenation of the shared features p^j , and the task-specific features from the previous layer $\hat{a}_i^{j-1}$:

$$Mask_u^l = h(g([p^l; f(\hat{e}_u^{(l-1)})])). \qquad (5)$$

Evaluation

We conduct extensive experiments on publicly available datasets. Results verify that:

- FedMAT improves performance for observed individuals;
- FedMAT helps with adaptation to new individuals.

Model	HHAR		PAMAP2		ExtraSensory		SmartJLU	
	Accuracy	macro-F1	Accuracy	macro-F1	Accuracy	macro-F1	Accuracy	macro-F1
DeepSense	94.12	93.43	89.37	90.67	65.62	64.17	84.71	80.56
AttenSense	94.22	94.98	88.11	88.31	67.26	66.82	85.09	82.11
DeepSense-MTL	96.45	96.08	91.37	90.43	70.98	71.19	87.37	83.01
AttenSense-MTL	96.15	95.93	90.10	90.32	71.75	71.03	87.10	**84.32**
Meta-HAR	96.02	95.85	90.47	89.92	72.32	71.29	86.40	80.13
FedMAT-noSMask	96.17	96.01	91.89	91.73	71.36	70.43	87.82	83.79
FedMAT-noFMask	95.29	94.62	90.14	90.25	69.12	69.09	82.14	78.25
FedMAT	**96.88**	**96.81**	**92.61**	**91.84**	**75.72**	**75.03**	**89.78**	83.02

Table 1: Overall comparison results on generalizing with existing individuals (unit:%).

Model	HHAR		PAMAP2		ExtraSensory		SmartJLU	
	Accuracy	macro-F1	Accuracy	macro-F1	Accuracy	macro-F1	Accuracy	macro-F1
DeepSense	91.13	90.88	80.01	78.51	60.22	58.53	76.91	74.14
AttenSense	90.41	90.22	81.53	82.11	64.12	60.17	78.67	74.05
DeepSense-MTL	91.02	91.46	84.31	85.31	63.18	58.13	79.09	76.53
AttenSense-MTL	92.81	91.98	82.72	83.12	62.15	59.03	80.04	74.58
Meta-HAR	93.13	92.82	**86.91**	85.41	68.16	62.92	82.04	80.45
FedMAT-noSMask	95.77	95.56	83.89	82.73	71.36	68.43	85.33	83.59
FedMAT-noFMask	93.89	93.62	86.04	85.65	69.12	66.09	82.12	80.50
FedMAT	**95.83**	**95.81**	86.72	**85.94**	**73.83**	**69.97**	**86.74**	**84.55**

Table 2: Overall comparison results on adapting to the new individuals (unit:%).

Evaluation

We conduct extensive experiments on publicly available datasets. Results verify that:

- Multi-task attention module learns heterogeneous features effectively;
- FedMAT adapts faster.

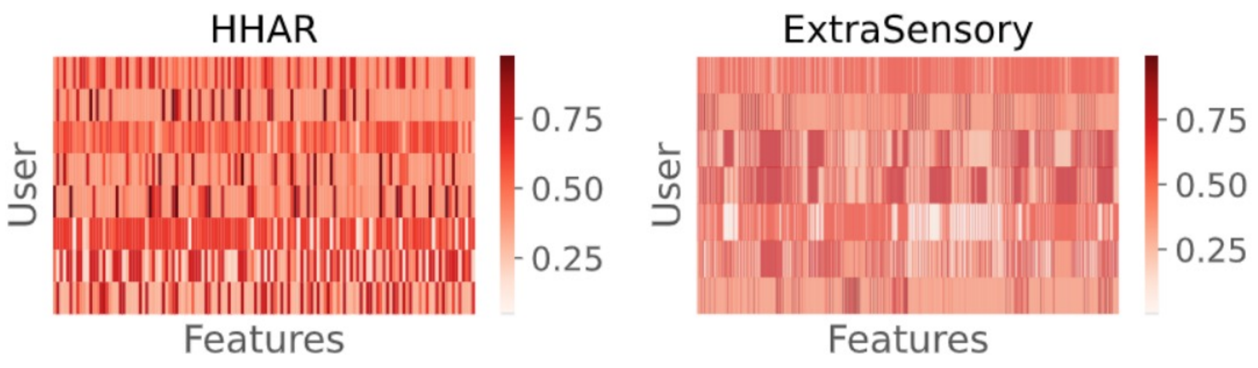

Figure 3: Visualization of attention-based mask on the HHAR and ExtraSensory datasets.

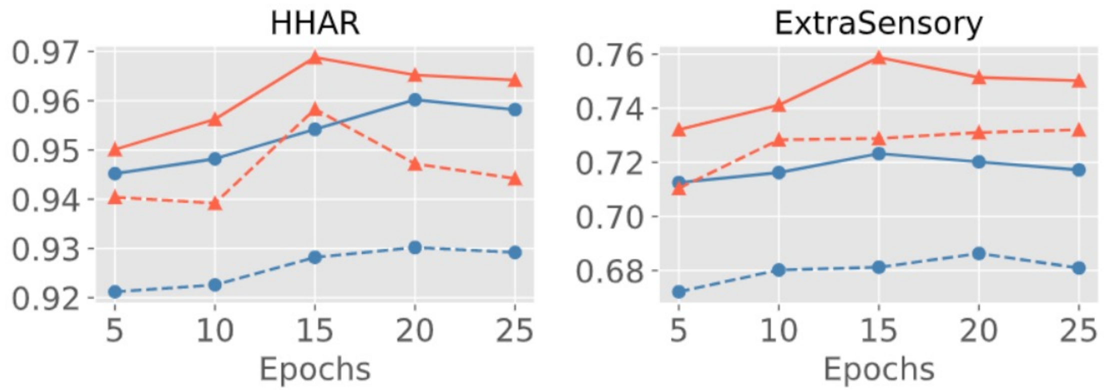

Figure 4: Evaluation of training epochs. Blue lines indicate performances of Meta-HAR and the red lines are for FedMAT. Dot lines refer to *macro-F_1* and plain lines are for *accuracy*

Thanks!

Qiang Shen[1], Haotian Feng[1], Rui Song[2], Stefano Teso[3], Fausto Giunchiglia[1,3] and Hao Xu [1,2]

1 College of Computer Science and Technology, Jilin University
2 School of Artificial Intelligence, Jilin University
3 University of Trento