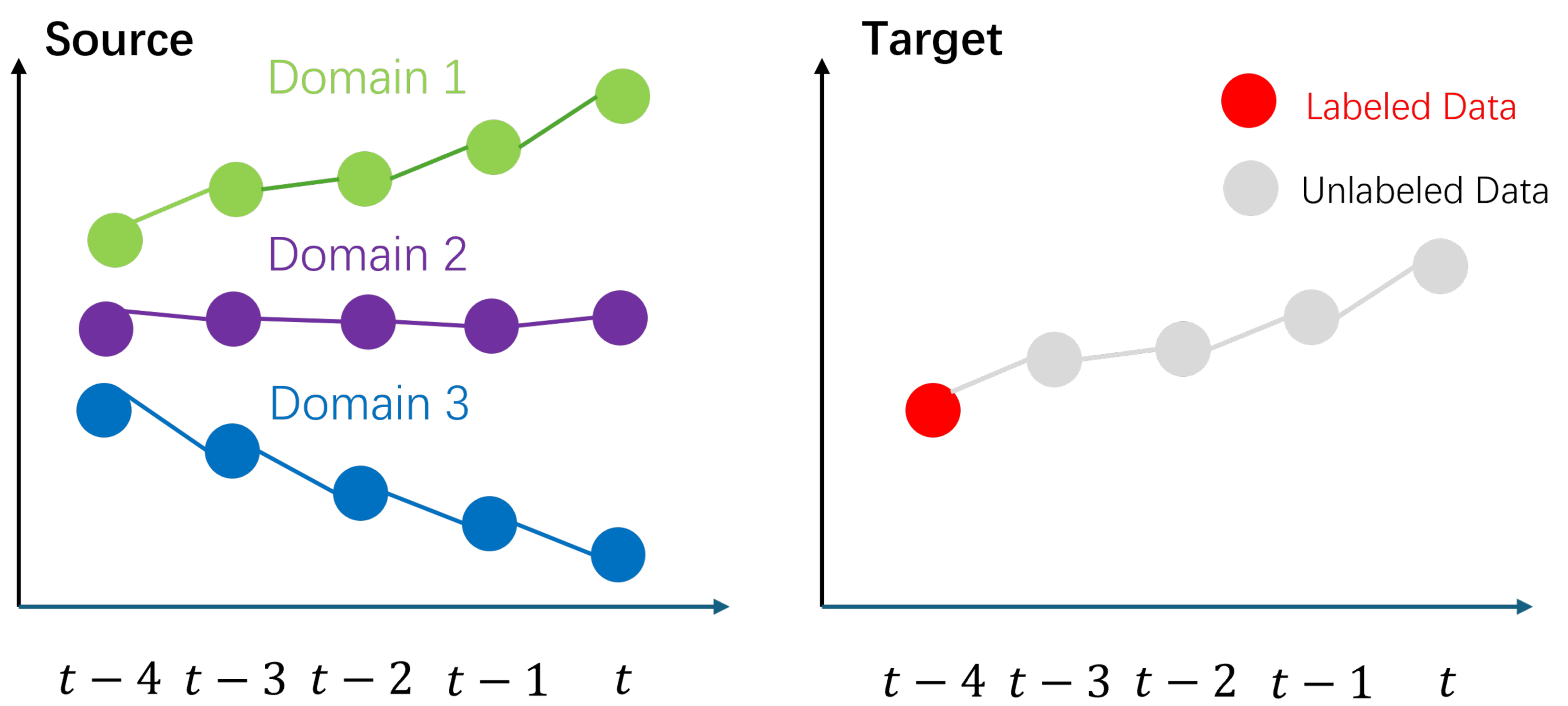
POND: Multi-Source Time Series Domain Adaptation with Information-Aware Prompt Tuning

NEC
NEC Laboratories America

Junxiang Wang*, Guangji Bai⁻, Wei Cheng*, Zhengzhang Chen*, Liang Zhao⁻, Haifeng Chen*

* NEC Laboratories America, ⁻ Emory University

Problem Background



Multi-source Time Series Domain Adaptation aims to leverage labeled data from multiple time series source domains to infer labels for unlabeled data in the target time series domains. In the example of the transponder fault classification problem, a model is trained under certain working modes (*e.g.*, single mode), and then it is applied to other working modes (*e.g.*, multimode).

Key Challenges: Existing methods primarily encode time series inputs from different source domains into domain-invariant representations, which often overlook domain-specific information such as global trends, local trends, and temporal patterns. It is difficult to learn dynamic domain-specific information and to evaluate whether the learned information reflects the true one.

Overview of PrOmpt-based domaiN Discrimination (POND)

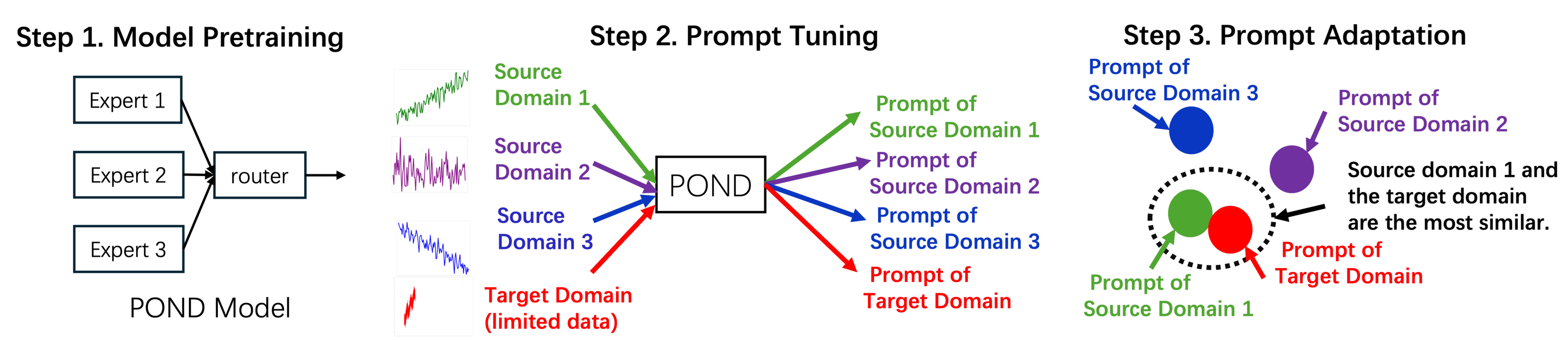


Figure 1. Pipeline of our proposed POND model

Our proposed POND model is implemented in three steps:

Model Pretraining: All experts and the router are pretrained by combining some labeled data from all source domains (*e.g.* 60%).

Prompt Tuning: Given the pretrained POND model, other labeled time series data from all source domains (*e.g.* 40%) are utilized to learn prompts from all source domains and the target domain.

Prompt Adaptation: The most similar source domain is selected by whose prompt is similar to the target domain, and this source domain is utilized for model prediction.

Capture Domain-Specific Information by Prompts

Prompts are utilized to capture domain-specific information by the labeled time series pair $(X_j^{(S_i)}, Y_j^{(S_i)})$. Let $P^{(S_i)}$ be the prompt of the source domain S_i, then, for the j-th time series input $X_j^{(S_i)}$, any time series model takes $[P^{(S_i)}, X_j^{(S_i)}]$ as its model input. We decompose $P^{(S_i)}$ into two components:

$$P^{(S_i)} = P + \Delta P^{(S_i)}$$

where P and $\Delta P^{(S_i)}$ are a common and domain-specific prompt respectively. $\Delta P^{(S_i)}$ captures dynamic information by a conditional module $g^{(S_i)}$ to generate instance-level prompts based on time series instances:

$$\Delta P_j^{(S_i)} = g^{(S_i)}(X_j^{(S_i)}; \zeta)$$

where ζ is a random variable, and the domain-level prompt $\Delta P^{(S_i)}$ is the aggregation of all instance-level prompts $\Delta P_j^{(S_i)}$ (*e.g.*, $\Delta P^{(S_i)} = \frac{1}{|S_i|}\sum_{j=1}^{|S_i|} \Delta P_j^{(S_i)}$).

Two Criteria for Good Prompts: Fidelity and Distinction

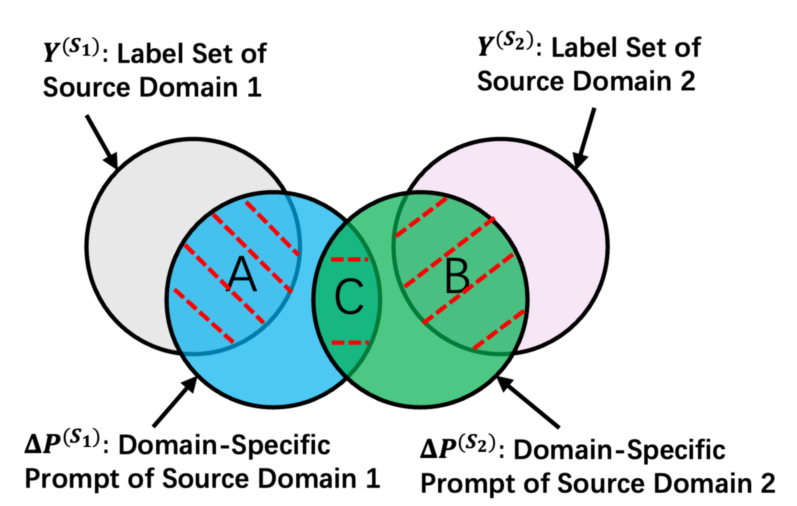


Figure 2. Illustration of two criteria: high fidelity and high distinction. Fidelity and distinction are represented as areas of $A + B$ and C, respectively.

High Fidelity requires that $\Delta P_j^{(S_i)}$ preserves the domain-specific information of the source domain S_i by minimizing the following fidelity loss:

$$\ell_F = \sum_{i=1}^{M} \sum_{j=1}^{|S_i|} Y_j^{(S_i)} \log f([\Delta P_j^{(S_i)}, X_j^{(S_i)}]).$$

where f is the POND model. We theoretically show that $\Delta P_j^{(S_i)}$, which minimizes the fidelity loss, possesses the following properties:

Property 1(Preserving Fidelity): The mutual information between $\Delta P_j^{(S_i)}$ and the label $Y_j^{(S_i)}$ is equivalent to that between the time series input $X_j^{(S_i)}$ and the label $Y_j^{(S_i)}$, *i.e.*, $MI(\Delta P_j^{(S_i)}, Y_j^{(S_i)}) = MI(X_j^{(S_i)}, Y_j^{(S_i)})$.

Property 2(Adding New Information): The generated prompt $\Delta P_j^{(S_i)}$ contains new information compared to the time series input $X_j^{(S_i)}$, *i.e.*, $H(\Delta P_j^{(S_i)}) \geq H(X_j^{(S_i)})$.

High Distinction requires that $\Delta P^{(S_i)}$ distinguishes the unique information of the source domain S_i from other source domains by minimizing the discrimination loss:

$$\ell_D = \sum_{i_1 \neq i_2} \mathbb{E} \log \frac{\exp(\mathrm{sim}(\Delta P^{(S_{i_1})}, \Delta P^{(S_{i_2})}))}{\sum_{i \neq i_1, i \neq i_2} \exp(\mathrm{sim}(\Delta P^{(S_{i_1})}, \Delta P^{(S_i)}))}.$$

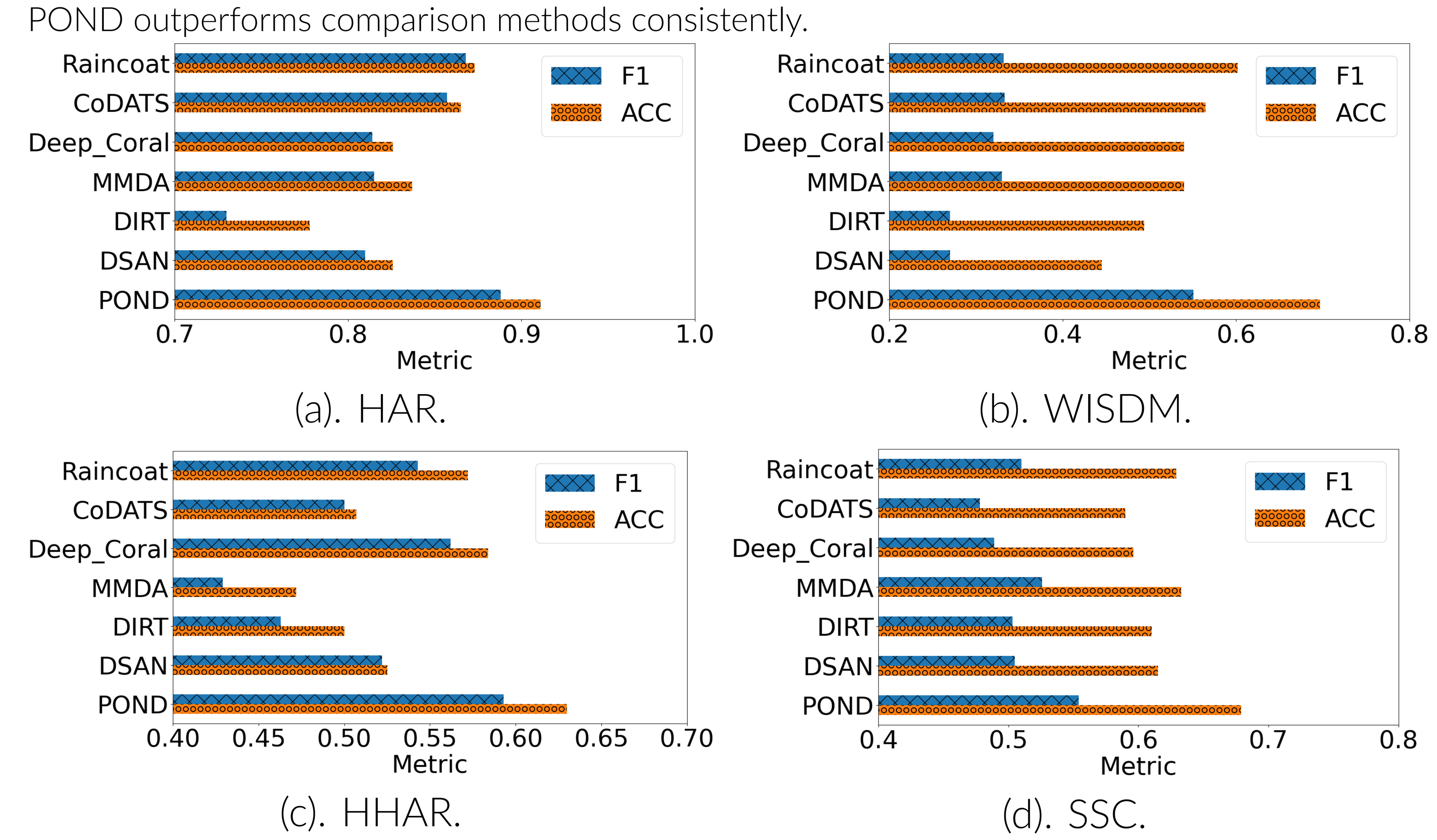
where $\Delta P^{(S_{i_1})}$ and $\Delta P^{(S_{i_2})}$ represent the domain-specific prompts of any two source domains S_{i_1} and S_{i_2}, and $\mathrm{sim}(\Delta P^{(S_{i_1})}, \Delta P^{(S_{i_2})})$ denotes the similarity score (*e.g.* cosine similarity) between two domain-specific prompts $\Delta P^{(S_{i_1})}$ and $\Delta P^{(S_{i_2})}$.

Experimental Verification

Table 1. F1-score on different scenarios of four datasets: the proposed POND model outperforms all comparison methods.

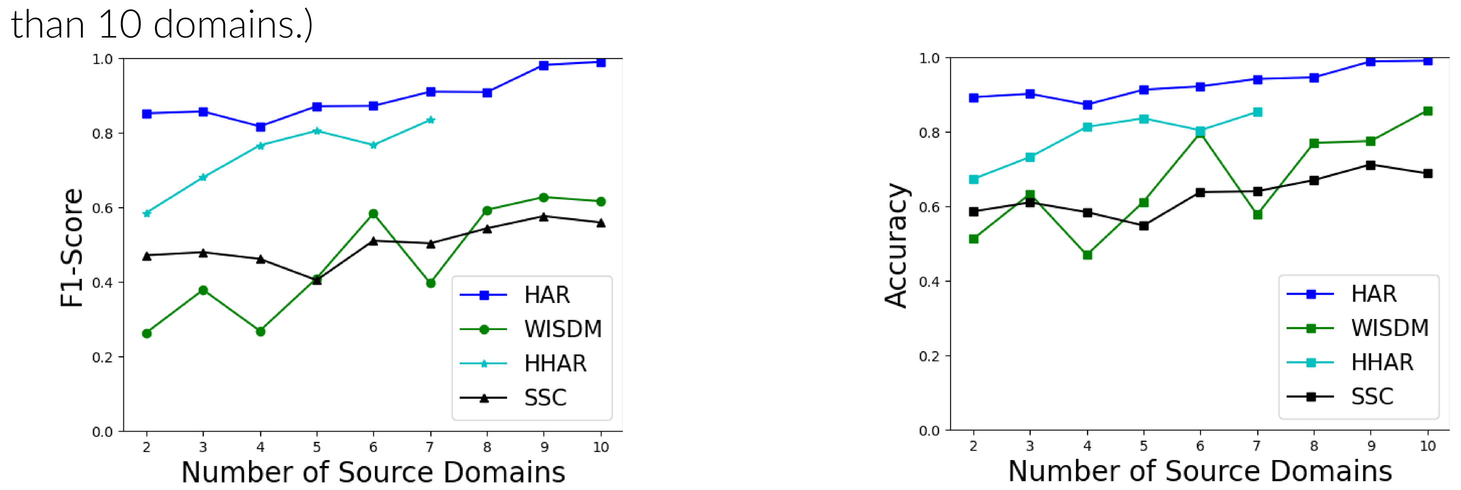
Scenario	Raincoat	CoDATs	Deep_Coral	MMDA	DIRT	DSAN	POND	Target Only
HAR 1-15 → 16	0.823 ± 0.094	0.767 ± 0.093	0.773 ± 0.082	0.679 ± 0.084	0.612 ± 0.135	0.738 ± 0.095	0.849 ± 0.021	0.856 ± 0.027
HAR 1-15 → 20	0.872 ± 0.142	0.932 ± 0.025	0.923 ± 0.023	0.921 ± 0.034	0.848 ± 0.101	0.929 ± 0.033	0.968 ± 0.021	0.983 ± 0.018
HAR 1-15 → 21	0.867 ± 0.141	0.903 ± 0.070	0.882 ± 0.028	0.974 ± 0.039	0.921 ± 0.090	0.909 ± 0.110	0.972 ± 0.021	1.000 ± 0.000
HAR 1-15 → 28	0.766 ± 0.107	0.775 ± 0.166	0.852 ± 0.044	0.778 ± 0.085	0.671 ± 0.175	0.783 ± 0.046	0.829 ± 0.018	0.853 ± 0.019
HAR 16-20 → 1	0.792 ± 0.072	0.744 ± 0.053	0.667 ± 0.077	0.654 ± 0.074	0.546 ± 0.060	0.698 ± 0.037	0.883 ± 0.017	0.986 ± 0.010
HAR 16-20 → 2	0.825 ± 0.048	0.821 ± 0.151	0.796 ± 0.055	0.651 ± 0.045	0.509 ± 0.050	0.652 ± 0.057	0.936 ± 0.017	0.943 ± 0.024
HAR 16-20 → 3	0.814 ± 0.028	0.746 ± 0.078	0.741 ± 0.058	0.657 ± 0.033	0.605 ± 0.056	0.565 ± 0.043	0.878 ± 0.018	0.978 ±0.013
HAR 16-20 → 4	0.679 ± 0.084	0.605 ± 0.082	0.479 ± 0.110	0.513 ± 0.058	0.336 ± 0.110	0.436 ± 0.032	0.754 ± 0.033	0.921 ± 0.018
WISDM 0-17 → 18	0.379 ± 0.061	0.384 ± 0.049	0.346 ± 0.023	0.297 ± 0.016	0.300 ± 0.041	0.287 ± 0.045	0.606 ± 0.020	0.705 ± 0.046
WISDM 0-17 → 20	0.354 ± 0.040	0.368 ± 0.039	0.376 ± 0.031	0.452 ± 0.098	0.347 ± 0.071	0.269 ± 0.064	0.570 ± 0.023	0.704 ± 0.051
WISDM 0-17 → 21	0.355 ± 0.057	0.310 ± 0.088	0.259 ± 0.018	0.250 ± 0.000	0.276 ± 0.055	0.245 ± 0.046	0.450 ± 0.026	0.636 ± 0.095
WISDM 0-17 → 23	0.306 ± 0.015	0.327 ± 0.075	0.318 ± 0.031	0.327 ± 0.023	0.271 ± 0.016	0.277 ± 0.044	0.482 ± 0.017	0.538 ± 0.034
WISDM 0-17 → 25	0.365 ± 0.030	0.540 ± 0.125	0.435 ± 0.043	0.436 ± 0.094	0.314 ± 0.107	0.353 ± 0.120	0.559 ± 0.050	0.672 ± 0.039
WISDM 0-17 → 28	0.399 ± 0.028	0.431 ± 0.033	0.418 ± 0.032	0.454 ± 0.064	0.304 ± 0.044	0.339 ± 0.030	0.656 ± 0.046	0.689 ± 0.048
WISDM 0-17 → 30	0.314 ± 0.020	0.305 ± 0.028	0.298 ± 0.023	0.359 ± 0.072	0.266 ± 0.035	0.246 ± 0.076	0.670 ± 0.039	0.791 ± 0.028
WISDM 18-23 → 5	0.648 ± 0.001	0.558 ± 0.129	0.534 ± 0.102	0.510 ± 0.020	0.549 ± 0.097	0.484 ± 0.055	0.652 ± 0.035	0.734 ± 0.095
WISDM 18-23 → 6	0.544 ± 0.074	0.565 ± 0.143	0.437 ± 0.078	0.543 ± 0.160	0.405 ± 0.089	0.454 ± 0.112	0.628 ± 0.033	0.872 ± 0.049
WISDM 18-23 → 7	0.588 ± 0.070	0.404 ± 0.117	0.530 ± 0.094	0.477 ± 0.060	0.518 ± 0.120	0.476 ± 0.127	0.672 ± 0.029	0.888 ± 0.035
HHAR 0-6 → 7	0.765 ± 0.142	0.652 ± 0.108	0.815 ± 0.105	0.641 ± 0.050	0.649 ± 0.005	0.730 ± 0.164	0.834 ± 0.014	0.861 ± 0.016
HHAR 5-8 → 2	0.321 ± 0.023	0.347 ± 0.082	0.309 ± 0.032	0.216 ± 0.032	0.276 ± 0.021	0.314 ± 0.095	0.352 ± 0.014	0.881 ± 0.018
SSC 0-9 → 16	0.578 ± 0.028	0.510 ± 0.044	0.537 ± 0.024	0.559 ± 0.027	0.523 ± 0.019	0.515 ± 0.044	0.568 ± 0.012	0.601 ± 0.018
SSC 0-9 → 17	0.511 ± 0.024	0.413 ± 0.118	0.452 ± 0.077	0.504 ± 0.060	0.530 ± 0.053	0.463 ± 0.081	0.559 ± 0.006	0.602 ± 0.014
SSC 0-9 → 18	0.605 ± 0.016	0.548 ± 0.037	0.544 ± 0.046	0.597 ± 0.032	0.574 ± 0.021	0.569 ± 0.046	0.604 ± 0.014	0.602 ± 0.013
SSC 0-9 → 19	0.562 ± 0.024	0.540 ± 0.052	0.531 ± 0.055	0.570 ± 0.044	0.565 ± 0.028	0.568 ± 0.080	0.570 ± 0.010	0.613 ± 0.019
SSC 10-12 → 8	0.294 ± 0.028	0.380 ± 0.066	0.379 ± 0.076	0.398 ± 0.060	0.322 ± 0.048	0.411 ± 0.046	0.470 ± 0.010	0.531 ± 0.019

Figure 3. The F1-score and accuracy of all methods on four benchmark datasets: the proposed POND outperforms comparison methods consistently.



(a). HAR. (b). WISDM. (c). HHAR. (d). SSC.

Figure 4. The F1-score and accuracy of the proposed POND model with different source domains: the performance grows with the increase of source domains. (The HHAR dataset has less than 10 domains.)



(a). F1-score. (b). Accuracy.