Highlights

Black-Box Time-Series Domain Adaptation via Cross-Prompt Foundation Models

M. T. Furqon, Mahardhika Pratama, Igor ŠKRJANC, Lin Liu, Habibullah Habibullah, Kutluyil Dogancay

- We propose the concept of CPFM for the BBTSDA problems. It is constructed under the time-series foundation model utilizing the prompt tuning strategy.
- We propose the idea of dual-branch network structure where each branch implements a unique prompt and complements each other. The final output is drawn from the aggregation of each branch output.
- We propose the idea of reconstruction learning in the prompt and input levels. The prompt reconstruction strategy creates distinct prompts generating complementary information while the input reconstruction method performs implicit domain alignment of the target domain by feeding the target domain samples without their labels.
- We numerically validate the advantage of CPFM using three datasets of different application domains. CPFM is capable of demonstrating the most encouraging performance outperforming SOTA algorithms with notable margins.

Black-Box Time-Series Domain Adaptation via Cross-Prompt Foundation Models⋆

M. T. Furqon[a], Mahardhika Pratama[a,*], Igor ŠKRJANC[b], Lin Liu[a], Habibullah Habibullah[a] and Kutluyil Dogancay[a]

[a]STEM, University of South Australia, Mawson Lakes Boulevard, Adelaide, 5095, Australia
[b]Faculty of Electrical and Computer Engineering, University of Ljubljana, Slovenia

ARTICLE INFO

Keywords:
transfer learning
domain adaptation
black-box domain adaptation
time-series

ABSTRACT

The black-box domain adaptation (BBDA) topic is developed to address the privacy and security issues where only an application programming interface (API) of the source model is available for domain adaptations. Although the BBDA topic has attracted growing research attentions, existing works mostly target the vision applications and are not directly applicable to the time-series applications possessing unique spatio-temporal characteristics. In addition, none of existing approaches have explored the strength of foundation model for black box time-series domain adaptation (BBTSDA). This paper proposes a concept of Cross-Prompt Foundation Model (CPFM) for the BBTSDA problems. CPFM is constructed under a dual branch network structure where each branch is equipped with a unique prompt to capture different characteristics of data distributions. In the domain adaptation phase, the reconstruction learning phase in the prompt and input levels is developed. All of which are built upon a time-series foundation model to overcome the spatio-temporal dynamic. Our rigorous experiments substantiate the advantage of CPFM achieving improved results with noticeable margins from its competitors in three time-series datasets of different application domains.

1. Introduction

The success of deep learning (DL) is largely attributed to the i.i.d conditions where training and testing samples follow the same distribution. However, this condition is too strict and does not mirror real-world situations where the training and deployment phases are often not the same, i.e., also known as the domain shift problem. Unsupervised domain adaptation (UDA) Ganin and Lempitsky (2014); Kang, Jiang, Yang and Hauptmann (2019); Furqon, Pratama, Liu, Habibullah and Doğançay (2024a) addresses this problem where the goal is to develop a model performing well on the unlabeled target domain given the labeled source domain under the presence of domain shifts between the source domain and target domain. Nonetheless, the classic UDA techniques require source-domain samples to be available, thus restricting their applications in the privacy and/or resource-constrained environments. Source-free domain adaptation (SFDA) approaches Liang, Hu and Feng (2020); Karim, Mithun, Rajvanshi, Chiu, Samarasekera and Rahnavard (2023); Litrico, Bue and Morerio (2023); Furqon, Pratama, Shiddiqi, Liu, Habibullah and Doğançay (2024b) address the drawback of the traditional UDA methods accessing only the source model rather than the source-domain samples. Nevertheless, the white box nature of the SFDA approaches does not fully protect the issue of privacy. The notion of black box domain adaptation (BBDA) Liang, Hu, Feng and He (2021); Yang, Peng, Wang, Zhu, Feng, Xie and You (2022) goes one step further than the SFDA where it only calls for application programming interface (API) for the target domain. Such domain adaptation is highly challenging because of the noisy pseudo label problem, i.e., the outputs of the source model obviously contain significant noise due to the domain shift problem whereas the discrepancies between two domains are not directly estimable because of the absence of source domain samples and source models. Fig 1 portrays the significant gap between the upper bound performance trained with the true class labels of the target domain and the lower bound performance delivered by the source model outputs in the human activity recognition (HAR) dataset. In a nutshell, there are many noisy pseudo labels generated by the source model. On the other hand, the BBDA problem is more challenging than the SFDA problem because of the absence of pre-trained parameters, thereby losing domain-specific information.

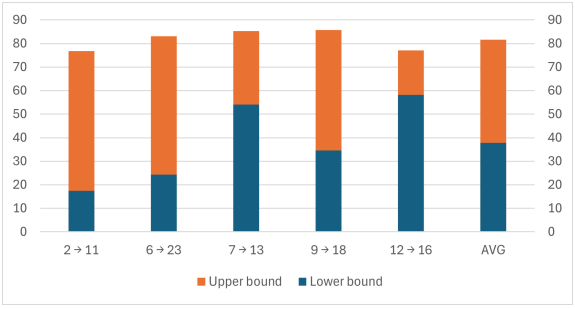

Figure 1: Lower bound vs Upper bound performance in the HAR dataset

The BBDA research topic has gained growing research attentions. In Liang et al. (2021), the so-called distill and

⋆

*Corresponding author
muhammad_tanzil.furqon@mymail.unisa.edu.au (M.T. Furqon); dhika.pratama@unisa.edu.au (M. Pratama); Igor.Skrjanc@fe.uni-lj.si (I. ŠKRJANC); lin.liu@unisa.edu.au (L. Liu); habibullah.habibullah@unisa.edu.au (H. Habibullah); kutluyil.dogancay@unisa.edu.au (K. Dogancay)
ORCID(s):

fine-tune (DINE) is proposed to address the problem of single and multi-source black-box domain adaptation and based on the idea of two-step knowledge adaptation. BiMem is proposed in Zhang, Huang, Jiang and Lu (2023) where it is built upon the concept of Atkinson-Shiffrin memory. BETA is designed in Yang et al. (2022) and puts forward knowledge distillation coupled with noisy-label learning. Co-MDA is put into perspective in Liu, Xi, Li, Xu, Bai and Zhao (2023) where the key idea lies in the concept of multi-domain attention and the co-learning framework. In Xiao, Ye, He, Li, Tang and Zhu (2024), the concept of dual experts is put into perspective to guide the domain adaptation process of two classifiers, while Jahan and Savakis (2023) proposes the curriculum learning strategy to avoid noisy pseudo labels. The problem of forgotten classes in BBDA is unveiled in Zhang, Shen, Lü and Zhang (2024). The idea of separation and alignment (SEAL) is proposed in Liu, Zhou, Ye and Li (2022); Xia, Zhao, Lyu, Huang, Hu, Chen and Wang (2024) where data samples are at first grouped into well-adapted and under-adapted regimes followed by the graph contrastive learning technique to improve model's representations. All these works merely focus on the vision application and are not directly applicable to the time-series domain possessing the spatio-temporal characteristic, i.e., their performance drops in the time-series problems.

To date, only two works in Ren and Cheng (2023); Jiao, Zhang, Li, Liu and Lin (2025) discuss the issue of black-box time-series domain adaptation (BBTSDA). Ren and Cheng (2023) is framed under the teacher-student learning with the temporal consistency loss. In addition, the shapley-enhanced method is incorporated to derive the contribution of each source domain. Jiao et al. (2025) applies the knowledge distillation concept followed by local and global regularization for fault diagnosis problems. Nevertheless, the issue of noisy labels remain unsolved as the big gap to the upper bound performance still exists. Beside, none of existing approaches explore the strength of the foundation model offering performance improvements. That is, the foundation model can be shared across different nodes while maintaining domain-specific prompts to address downstream tasks. Note that, in realm of BBDA, the prompts cannot be exchanged. They are kept private to preserve the privacy issue.

This paper proposes Cross Prompt Foundation Models (CPFM) for the BBTSDA problems. The domain adaptation stage is developed using the idea of reconstruction learning in both input and prompt level. The input reconstruction approach functions as an implicit domain adaptation of the target domain while the prompt reconstruction strategy assures distinct prompt making possible complementary information to be explored by the dual-branch network structure. The CPFM is built upon a dual-branch network structure where distinct prompts are mounted in each branch while their outputs are combined in such a way that complementary information is fused. That is, we implement the notion of prompt tuning Wang, Zhang, Ebrahimi, Sun, Zhang, Lee, Ren, Su, Perot, Dy and Pfister (2022) where the backbone model is frozen leaving only small trainable parameters, called prompts, to be tuned. All of which are built upon a time-series foundation model Goswami, Szafer, Choudhry, Cai, Li and Dubrawski (2024) pretrained with abundant time-series datasets guaranteeing decent model's generalizations under spatio-temporal time-series problems. Our major contributions are listed:

- We propose the concept of CPFM for the BBTSDA problems. It is constructed under the time-series foundation model utilizing the prompt tuning strategy.
- We propose the idea of dual-branch network structure where each branch implements a unique prompt and complements each other. The final output is drawn from the aggregation of each branch output.
- We propose the idea of reconstruction learning in the prompt and input levels. The prompt reconstruction strategy creates distinct prompts generating complementary information while the input reconstruction method performs implicit domain alignment of the target domain by modeling the target domain samples without their labels.
- We numerically validate the advantage of CPFM using three datasets of different application domains. CPFM is capable of demonstrating the most encouraging performance outperforming SOTA algorithms with noticeable margins.

2. Related Works

2.1. Time-Series Domain Adaptation

Time-Series Domain Adaptation (TSDA) has been studied where the goal is to overcome the temporal nature of time-series data which does not exist in the vision application Liu et al. (2023) in addition to that of the domain shifts between the source domain and the target domain. There exist two approaches in this domain: adversarial-based approach and discrepancy-based approach. AdvSKM Liu et al. (2023) constitutes a discrepancy-based approach using the MMD approach coupled with the spectral kernel method to minimize the domain gap between the source domain and the target domain and to take into account the temporal dependencies of the time-series samples. The association structure is designed in SASA for TSDA Cai, Chen, Li, Chen, Zhang, Ye, Li, Yang and Zhang (2020). MDAN Furqon et al. (2024a) puts forward the idea of intermediate domain to dampen the discrepancies between the source domain and the target domain. On the other hand, the adversarial-based approach utilizes a domain discriminator to play the adversarial game reducing the domain gap. CoDATS Wilson, Doppa and Cook (2020) implements such concepts for TSDA in the human activity recognition problems. DAATTN combines the adversarial learning with the attention sharing mechanism Jin, Park, Maddix, Wang and Yan (2021). SLARDA Ragab, Eldele, Chen, Wu, Kwoh and Li (2021) presents an autoregressive domain discriminator for the adversarial training approach. Notwithstanding that

these approaches have been successful for TSDA, they call for source-domain samples and a pretrained source model to be shared during the domain adaptation step, thus raising the privacy and storage concerns.

2.2. Source-Free Domain Adaptation

The issue of privacy has led to the advent of source-free domain adaptation (SFDA) where the goal is to generalize well over the unlabeled target domain with the absence of source domain samples. Only a pretrained source model is shared for domain adaptation. Liang et al. (2020) relies on the self-training mechanism with the cluster's structure and Li, Jiao, Cao, Wong and Wu (2020) utilizes the generative model to address the absence of source domain samples. Chen, Wang, Darrell and Ebrahimi (2022a) puts forward the idea of self-supervised learning while Karim et al. (2023) also uses the self-supervised learning technique combined with the notion of curriculum learning to prevent early memorization of noisy pseudo labels. The concept of loss re-weighting using the entropy estimation is put forward in Litrico et al. (2023). The aforementioned methods are designed for vision applications excluding any spatio-temporal properties. The proposals of source-free time-series domain adaptation (SFTSDA) are exemplified in Zhao, Feng, Li, Song, Liang and Chen (2023) using the GMM concept for seizure predictions and Ragab, Eldele, Wu, Foo, Li and Chen (2023) integrating the time-series imputation strategy. In Furqon et al. (2024b), the time-frequency concept is introduced for SFTSDA. Nonetheless, the SFDA concept does not fully protect the client's privacy because the source model is not shareable in many applications. That is, source-domain samples can be reconstructed by certain techniques such as the deepinversion due the presence of the pretrained source model.

2.3. Black-Box Domain Adaptation

Black-Box Domain Adaptation (BBDA) goes one step ahead of SFDA where only the API of the source model is offered for domain adaptations. That is, one can only elicit soft or hard labels of the source model for further preserving the client's privacy. Liang et al. (2021) proposes the so-called DINE for single and multi-source BBDA using the two steps knowledge adaptation. The concept of Atkinson-Shiffrin memory is realized in Zhang et al. (2023) for BBDA. The combination of multi-domain attention and co-learning is proposed in Co-MDA Liu et al. (2023) for BBDA while BETA is devised in Yang et al. (2022) using the concept of knowledge distillation and noisy-label learning. The concept of dual experts is proposed in Xiao et al. (2024) while the curriculum learning approach is put forward in Jahan and Savakis (2023). The issue of forgotten classes in BBDA is discussed in Zhang et al. (2024). The separation and alignment (SEAL) method is put into perspective in Liu et al. (2022); Xia et al. (2024) and achieves SOTA results in the vision applications. All these methods are designed for vision application and are not readily applicable for the time-series applications. To the best of our knowledge, the black-box time-series domain adaptation (BBTSDA) problem is only addressed in Ren and Cheng (2023); Jiao et al. (2025) where the temporal consistency loss and the shapley-enhanced method are integrated in Ren and Cheng (2023) while Jiao et al. (2025) presents the concept of knowledge distillation followed by local and global regularization. None of existing methods explore the advantage of foundation models possibly offering promising alternatives. That is, the foundation model can be kept fixed and shared across each node while only performing parameter efficient fine-tuning strategies for domain-specific problems. In other words, the foundation model captures general spatio-temporal characteristics of time-series data because it is pre-trained using massive time-series problems. To protect the privacy issue, the prompts providing domain-specific information can be kept private for each client.

3. Preliminaries

3.1. Problem Definition

Given a target model $f_{\phi_t}(g_{\psi_t}(.))$ where $g_{\psi_t}(.) : \mathcal{X} \rightarrow \mathcal{Z}$ is a feature extractor mapping the input space to the latent space and $f_{\phi_t}(.) : \mathcal{Z} \rightarrow \mathcal{Y}$ is a projector converting the latent space to the label space, the goal of black-box time-series domain adaptation (BBTSDA) is to perform well on an unlabelled target domain $\mathcal{D}_T$ having N_t unlabelled samples $\{x_i\}_{i=1}^{N_t}$ where $x_i \in \mathfrak{R}^{T \times D}$. T, D respectively stand for the length of a time-series sample and the number of variables. The domain adaptation process is guided by M black-box predictors $\{f_{\phi_{s_i}}(g_{\psi_{s_i}}(.))\}_{i=1}^{M}$ pre-trained by an $i-th$ source domain $\mathcal{D}_{S_i}, i \in \{1, ..., M\}$ consisting of N_{s_i} labelled samples $\{(x_j, y_j)\}_{j=1}^{N_{s_i}}$. The underlying challenge is perceived in the issue of domain shifts where the target domain and each source domain follow an independent distribution such that $\mathcal{D}_T \neq \mathcal{D}_{S_i} \neq \mathcal{D}_{S_j}, i, j \in \{1, ..., M\}$. In addition, the BBTSDA fully preserves the client's privacy where the source-domain samples $\{(x_j, y_j)\}_{j=1}^{N_{s_i}}$ and the parameters of the black box predictors $\{\psi_{s_i}, \phi_{s_i}\}_{i=1}^{M}, i \in \{1, ..., M\}$ are unavailable for domain adaptations. That is, it is only navigated by the API of the black box predictors for the target domain generating the soft-label $\hat{y}_i^t = f_{\phi_{s_i}}(g_{\psi_{s_i}}(x^t))$ or the hard label $\hat{C}_i^t = \arg\max_c \delta(f_{\phi_{s_i}}(g_{\psi_{s_i}}(x^t)))$ where $\delta(.)$ is a softmax function. Since we exploit the foundation model in this paper, the prompts of the foundation models of the source domains are kept private. We limit our discussion in the closed-set scenario where the source domain and the target domain share the same label space $\mathcal{Y}_s = \mathcal{Y}_t$.

3.2. Foundation Model

CPFM is built upon MOMENT Goswami et al. (2024), a family of open-source foundation models for general-purpose time-series analysis. MOMENT is pre-trained using abundant time-series datasets in the reconstruction fashion. First, a time-series is broken-down into N dis-joint sub-sequences with a length of P termed, patches. Each patch is

projected into D dimensional embedding using a trainable linear projector if unmasked or a designated learnable mask embedding if masked. These N patch embedding becomes an input to the transformer model and maintains its shape $(1 \times D)$. It is used to reconstruct the masked and unmasked time-series samples using a lightweight projection head. The transformer encoder follows a modification of the original transformer Raffel, Shazeer, Roberts, Lee, Narang, Matena, Zhou, Li and Liu (2019) where the additive bias of the layer norm is removed and placed before the residual connection. It uses relational positional embedding scheme. No decoder is applied to allow the architectural modifications for task-specific fine-tuning.

3.3. Prompt Tuning

The prompt tuning concept Wang et al. (2022) is implemented in the CPFM where a small-sized external parameter, namely prompt, is injected in the multi-head self-attention layer (MSA) of the foundation model leaving the backbone network frozen, thus significantly decreasing the number of trainable parameters. That is, the learning process is only localized to the prompts. The prompt tuning technique modifies the input to the MSA layer. Let $p \in \Re^{L_p \times D}$ be the prompt and $h_Q, h_K, h_V \in \Re^{L \times D}$ be the input query, key and value, the prompt tuning method prepends the prompts to the input token which is equivalent to concatenate the same prompt parameter to the input query, key and values Wang et al. (2022) $MSA([p; h_Q], [p; h_k], [p; hv])$ where $[;]$ stands for the concatenation operation along the sequence length dimension. This leads to the increase of the output length $\Re^{(L+Lp)\times D}$.

4. Cross Prompt Foundation Model (CPFM)

Our algorithm, namely CPFM, is constructed under the time-series foundation model Goswami et al. (2024) where the idea of prompt tuning Wang et al. (2022) is integrated to adapt to downstream tasks. The dual-branch network structure is devised where each branch is inserted with unique prompts to explore different aspects of data distributions. Their outputs are aggregated to deliver final outputs and produce complementary information. The domain adaptation phase adopts the idea of reconstruction learning in the prompt level and the input level. The prompt reconstruction method is meant to generate distinct prompts offering complementary information while the input reconstruction approach performs the domain alignment step where input samples of the target domain are reconstructed. Fig. 2 shows the workflow of our approach.

4.1. Dual Branch Network Structure

CPFM is underpinned by the idea of cross-prompts working collaboratively to reject noisy pseudo-labels. This concept implements dual prompts for every source domain: p_i^1 and p_i^2 denoting respectively the first and second prompts in respect to the $i-th$ source domain while relying on the same foundation model frozen during the training process. Hence, this strategy minimizes the memory burdens because the same foundation model is applied to each source domain $\{\psi_t\} = \{\psi_{s_i}\}, i \in \{1, ..., M\}$. That is, they differ from each other only in the use of different prompts $p_i^1 \neq p_i^2 \neq p_j^1 \neq p_j^2, i, j\{1, ..., M\}, i \neq j$. In other words, their prompts are initialized differently. The embedding of the foundation model for the $i-th$ source domain can be expressed:

$$g_\psi(x; p_i^1) = MSA([p_i^1; Q_{j,k}], [p_i^1; K_{j,k}], [p_i^1; V_{k,j}]) \quad (1)$$

$$g_\psi(x; p_i^2) = MSA([p_i^2; Q_{j,k}], [p_i^2; K_{j,k}], [p_i^2; V_{k,j}]) \quad (2)$$

where $Q_{j,k}, K_{j,k}, V_{j,k}$ stand for the query, key and value of the $j-th$ head of MSA layer of the $k-th$ encoder. That is, the prompts are injected in every MSA head of the encoder layer. This mechanism results in different predictions due to the use of different linear heads fine-tuned during the training process $\{\phi_t^1\} \neq \{\phi_t^2\} \neq \{\phi_{s_i}\}, i \in \{1, ..., M\}$. Suppose that $o_i^1 = \sigma(f_{\phi_i^1}(g_\psi(x; p_i^1)))$ and $o_i^2 = \sigma(f_{\phi_i^2}(g_\psi(x; p_i^2)))$, the final output of the two networks are aggregated Wang, Yang, Tan, Bai and Zhou (2023) as follows:

$$o_i = \alpha o_i^1 + \beta o_i^2 \quad (3)$$

where $\alpha = \frac{\max_{c\in[0,C]} o_i^1}{\max_{c\in[0,C]} o_i^1 + \max_{c\in[0,C]} o_i^2}$ and $\beta = \frac{\max_{c\in[0,C]} o_i^2}{\max_{c\in[0,C]} o_i^1 + \max_{c\in[0,C]} o_i^2}$. $\sigma(.)$ is a softmax function. Unlike Han, Yao, Yu, Niu, Xu, Hu, Tsang and Sugiyama (2018); Liu et al. (2023) maintaining the predictions of two networks, we only retain the aggregations of the two predictions for simplicity. *Note that notwithstanding that the backbone networks are the same across the source and target domains, the prompts of the source domain are kept private and not to be shared during domain adaptations.* p_i^1, p_i^2 *are the prompts of the target domain when using the* $i-th$ *source model as the teacher.*

4.2. Domain Adaptation Phase

4.2.1. Prompt Reconstruction

Since the prompts are high-dimensional and required to reject different types of errors in the collaborative learning mechanism, they need to be distinct and do not contain redundant information $p_i^1 \neq p_i^2$ and across all source domains. To this end, we are inspired by Chen, Wu and Jiang (2022b) that high-dimensional data usually lies in a lower dimensional manifold and thus auto-encoders can be applied for improved alignments. The goal is to learn a domain-invariant latent subspace of denoised prompts where redundant information is removed via the reconstruction of the learned prompts. Suppose that $\hat{h}(.)$ denotes a projection function mapping the prompts $p_i^{1,2}$ into a lower dimensional manifold and $\tilde{h}(.)$ stands for a back-projection function projecting the vectors back into soft prompts $\hat{p}_i^{1,2}$. We follow the same architecture as Chen et al. (2022b) where $\hat{h}(.)$ is implemented as a one-layer feed-forward network while $\tilde{h}(.)$ is a two-layer nonlinear perceptron as follows:

$$\hat{h}(p_i^{1,2}) = W_1 p_i^{1,2} + b_1 \quad (4)$$

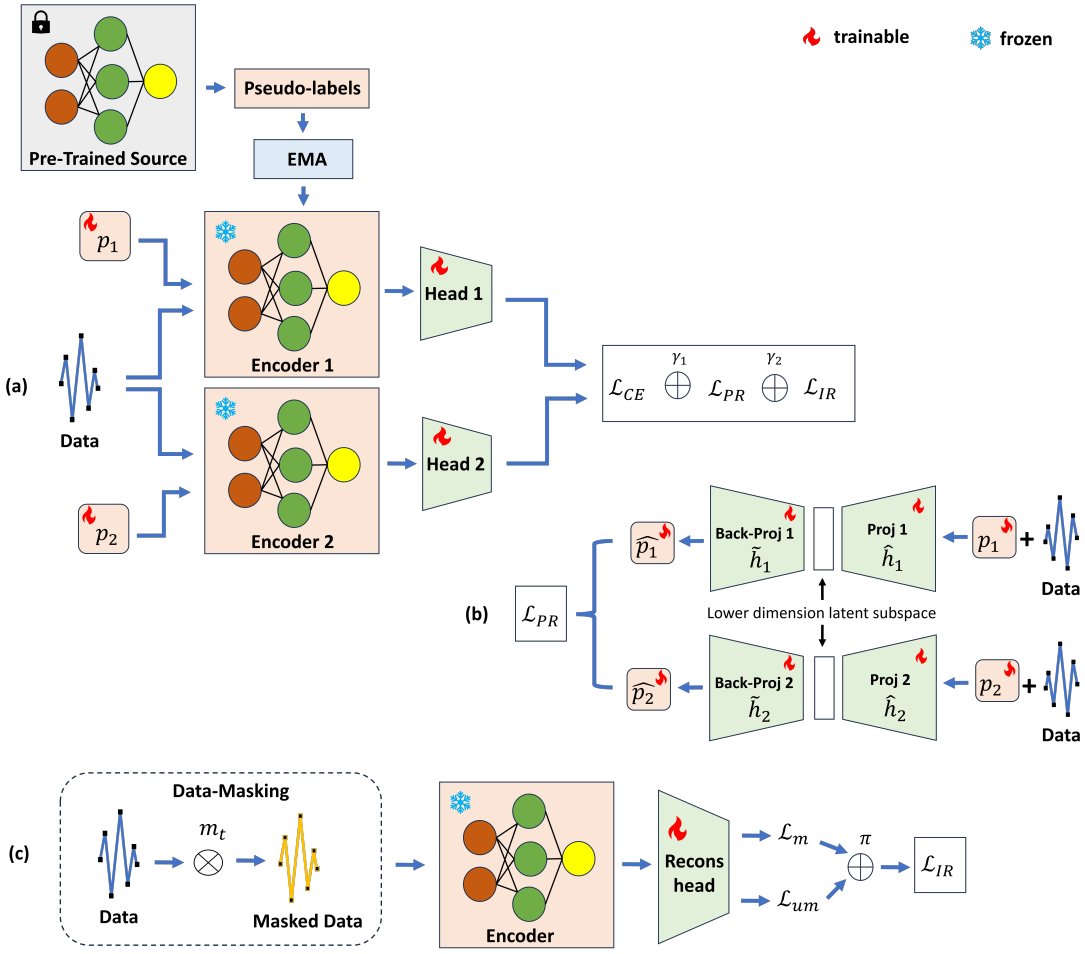

Figure 2: (a) CPFM learning policy where only the prompts and classification heads are tuned leaving the backbone network frozen; (b) Prompt Reconstruction is devised to deliver distinct prompts in each network branch; (c) Input Reconstruction learns the structure of the target domain without any label information.

$$\tilde{h}(\hat{h}) = W_3 tanh(W_2\hat{h} + b_2) + b_3 \tag{5}$$

The objective is to minimize the prompt reconstruction loss:

$$\mathcal{L}_{PR} = \frac{1}{2M}\sum_{i=1}^{M}\sum_{j=1}^{2}||\hat{p}_i^j - p_i^j||_2^2 \tag{6}$$

4.2.2. Input Reconstruction

Given the unlabelled target domain $\mathcal{D}_T$, the input reconstruction process is performed to the target domain samples x_t Furqon et al. (2024a). That is, a binary mask m_t is generated where x_t is zeroed if $m_t = 1$. The input reconstruction loss comprises masked and unmasked components as follows

$$\mathcal{L}_m = \frac{1}{D\sum_{i=1}^{N_t} m_t}\sum_{i=1}^{N_t} m_t||\delta(f_{\phi_t}(g_{\psi_t}(x_t))) - x_t||_2^2 \tag{7}$$

$$\mathcal{L}_{um} = \frac{1}{D(S - \sum_{i=1}^{N_t} m_t)} \sum_{i=1}^{N_t}(1 - m_t)||\delta(f_{\phi_t}(g_{\psi_t}(x_t))) - x_t||_2^2 \tag{8}$$

where $\delta(.)$ is a projector. The input reconstruction loss is defined as a combination of the masked and unmasked components.

$$\mathcal{L}_{IR} = \pi\mathcal{L}_m + (1 - \pi)\mathcal{L}_{um} \tag{9}$$

where $\pi \in [0, 1]$ is a trade-off constant. The input reconstruction strategy functions as an implicit domain alignment phase since it learns the structure of the target domain.

The overall objective function is mathematically expressed:

$$\mathcal{L} = \mathcal{L}_{CE} + \gamma_1\mathcal{L}_{PR} + \gamma_2\mathcal{L}_{IR} \tag{10}$$

where $\gamma_{1,2}$ are a trade-off constant controlling the strength of the prompt reconstruction loss and the input reconstruction

loss. $\mathcal{L}_{CE}$ stands for the cross-entropy loss computed using the pseudo-labels information or the soft label of the $i-th$ source model $\hat{y}_i^t = \sigma(f_{\phi_{s_i}}(g_{\psi_{s_i}}(x^t; p_i^{s_i})))$ where $p_i^{s_i}$ is the corresponding prompt. Nonetheless, the soft label likely contains noises leading to poor generalizations. We apply the exponential moving average (EMA) rule to alleviate the noisy pseudo label problem Jiao et al. (2025); Xiao et al. (2024).

$$\hat{y}_i^t = \gamma \hat{y}_i^t + (1-\gamma)\hat{y}^t \tag{11}$$

where $\hat{y}^t$ stands for the prediction of the target model and γ denotes the trade-off parameters simply set to 0.7. In other words, the prediction of the $i-th$ source model is slowly forgotten and replaced by the prediction of the target model overtime. In the first epoch, the output of the source model is computed as per the following equation to reduce inaccurate and even incomplete predictions.

$$\hat{y}_i^t = \begin{cases} \hat{y}_i^t, & Top^1 \\ \frac{(1-\hat{y}_i^t)}{K-1}, & otherwise \end{cases} \tag{12}$$

where Top^1 is the top 1 index of the prediction in $\hat{y}_i^t$.

4.3. Multi-Source Domains Case

In realm of multi-source domain adaptation, CPFM is driven by M teachers representing M different but related source domains and delivering predictions of unlabelled target-domain samples. The final output is mathematically expressed as follows:

$$\hat{y}^t = \sum_i^M \lambda_i \hat{y}_i^t \tag{13}$$

where $\hat{y}_i^t = f_{\phi_{s_i}}(g_{\psi_{s_i}}(x^t; p_i^{s_i}))$ and $\lambda_i \in [0,1]$ is a weighting coefficient of the $i-th$ source domain also known as the transferability weight. The weight determines the influence of the $i-th$ source domain toward the knowledge transfer.

Because of the absence of source domain samples, the transferrability weight $\{\lambda_i\}_{i=1}^M$ is estimated by the difficulty of knowledge transfer reflected by the prediction's uncertainty Lü, Kang and Li (2024).

$$\eta_i = \frac{1}{\mathcal{H}(\sigma(f_{\phi_{s_i}}(g_{\psi_{s_i}}(x^t; p_i^{s_i}))))} = \frac{1}{\mathcal{H}(\hat{y}_i^t)} \tag{14}$$

where $\mathcal{H}(.)$ denotes the Shannon entropy. That is, we consider the soft outputs of the $i-th$ source model. The shannon entropy is inversely proportional to the prediction confidence where the higher the entropy the less confidence the model is. This implies similarity between the $i-th$ source domain and the target domain where a low uncertainty is seen as a high similarity between the two domains. The final transferrability weight is enumerated from the inverse of Shannon entropy. The transferrability weight is normalized.

$$\lambda_i = \frac{\eta_i}{\max_{i=1,\dots,M}(\eta_i)} \tag{15}$$

At each epoch, the normalized transferrability weight is updated using a moving average formula.

$$\lambda_i^e = \alpha \lambda_i^{e-1} + (1-\alpha)\lambda_i^e \tag{16}$$

where $\alpha = \frac{N_p}{N_T}$ is a scaling coefficient, while N_p, N_T respectively denote the number of pseudo labels and target-domain samples.

4.4. Algorithm

The learning policy of CPFM is outlined in Algorithm 1 where, at first, the foundation models trained across diverse time-series datasets are loaded. The backbone network is frozen to enjoy generalized features of the foundation model, while the domain adaptation phase is done by tuning only the prompts. We apply the dual-branch network architecture meaning that two different prompts initialize our model and the final output is aggregated as per (3). The source model is trained in the source domain. Once completed, they are set as the teacher models for the target domain and their outputs are obtained as per the EMA formula (11) where their influence decay and is taken over by the target model as the training process runs. Once eliciting the teacher output, this knowledge is distilled to the target model by the cross entropy loss. We also calculate the input reconstruction loss as an implicit domain adaptation loss and the prompt reconstruction loss to avoid redundant prompts. Finally, the total loss is calculated as per (10) and induces the parameter learning process.

5. Complexity Analysis

Following the pseudo-code in Algorithm 1, CPFM has several operations e.g. obtain source model soft-label, initialize or update teacher buffer, obtain target model logits, perform input reconstruction, perform prompt reconstruction, calculate model losses, and update model parameters. Suppose that N is the total number of samples in the target dataset consisting of the number B of batches that satisfies $\sum_{b=1}^B N_b = N$, E is the number of training epochs, R is the size of memory buffer where $R < N$ and M is the number of teacher models. Let C denote the complexity of a process, the complexity of the proposed method can be written:

$$\begin{aligned} C(CPFM) = {} & C(ObtainTeacherSoftLabel) \\ & + C(InitializeUpdateTeacherBuffer) \\ & + C(ObtainTargetLogits) \\ & + C(InputReconstruction) \\ & + C(PromptReconstruction) \\ & + C(\mathcal{L}_{CE}) + C(\mathcal{L}_{IR}) + C(\mathcal{L}_{PR}) \\ & + C(UpdateModelParameters) \end{aligned} \tag{17}$$

Table 1
Dataset characteristics (Ch: # channels, K: # classes, S: sample length)

Dataset	Ch	K	S	# Training	# Testing
MFD	1	3	5120	7312	3604
UCIHAR	9	6	128	2300	990
SSC	1	5	3000	14280	6130

$$\begin{aligned} C(CPFM) = E.\sum_{b=1}^{B} N_b(O(M) + O(M.R) + O(1) \\ +O(1) + O(1) + O(1) + O(1) \\ +O(1) + O(1)) \end{aligned} \tag{18}$$

$$C(CPFM) = O(E.M.\sum_{b=1}^{B} N_b) + O(E.M.R.\sum_{b=1}^{B} N_b) \tag{19}$$

since $\sum_{b=1}^{B} N_b = N$ and M is a small number ($M < 10$), then the complexity of CPFM can be written as:

$$\begin{aligned} C(CPFM) = O(E.M.N) + O(E.M.R.N) \\ C(CPFM) = O(E.R.N) \end{aligned} \tag{20}$$

6. Experiments

6.1. Datasets

The advantage of our method, CPFM, is rigorously evaluated with three datasets of different application domains: human activity recognition, machine fault diagnosis and sleep stage classifications. The characteristics of the three datasets are summed up in Table 1

HAR dataset constitutes a human activity recognition dataset using three sensors to monitor three dimensional body movements leading to 9 channels per sample. We follow the same configuration of Ragab et al. (2023) where five cross-users experiments are set up. That is, a model is developed using a dataset of one user and subsequently evaluated with another user dataset.

SSC dataset is an EEG dataset monitoring sleep stages of five different classes. We adopt the same dataset as Ragab et al. (2023) using the sleep EDF dataset. We utilize a single channel, namely Fpz-Cz and 5 cross-users experiments from 10 subjects are set up.

MFD dataset is a bearing fault diagnosis problem initiated by the university of Paderborn. The fault is detected using the vibration signal and this dataset comprises four working conditions where each condition represents one domain. As with Ragab et al. (2023), the five cross-conditions experiments are put forward to evaluate the consolidated algorithms.

6.2. Baseline Algorithms

CPFM is compared with five state-of-the art black box domain adaptation methods: DINE Liang et al. (2021), CoMDA Liu et al. (2023), BETA Yang et al. (2022), SEAL Xia et al. (2024), RFC Zhang et al. (2024). All consolidated algorithms are executed under the same computational environments, i.e., 2 NVIDIA A5000 GPU with 24 GB of RAM, using their official implementations. CPFM is developed under the pytorch library and its source code is made publicly available in `https://github.com/furqon3009/CPFM` All algorithms are configured under the same architecture as CPFM where the moment foundation model Goswami et al. (2024) is used using the notion of prompt tuning Wang et al. (2022) to ensure fair comparisons. Because of limited computational resources, we are only able to run baseline algorithms with one random seed. This is mainly due to

Algorithm 1 CPFM

Input: Source model $f_{\phi_{s_i}}(g_\psi(\cdot))$, target models $f_{\phi_{t_i}^1}(g_\psi(\cdot))$, $f_{\phi_{t_i}^2}(g_\psi(\cdot))$, linear head parameters ϕ_t^1, ϕ_t^2, prompts p_i^1, p_i^2, prompting function f_{prompt}, target dataset $D_T = \{x_i\}_{i=1}^{N_T}$, number of samples N_T, number of epochs E, number of batches B, number of models f, number of source/teacher models M, teacher buffer R
Output: Configuration of CPFM
Procedure:
Load the foundation model g_ψ
Initialize p_i^1, p_i^2
Generate prompted architecture $g_\psi(x_i; p_i)$ ▷ Attach prompts to MSA layers via f_{prompt}
Generate R
for $m = 1$ to M **do**
    Forward through the teacher models $f_{\phi_{s_i}}(g_\psi(\cdot))$, obtain the soft-labels $\hat{y}_i^t$ as per (13)
end for
for $e = 1$ to E **do**
    for $b = 1$ to B **do**
        for $f = 1$ to 2 **do**
            for $r = 1$ to R **do**
                Initialize or buffer with $\hat{y}_i^t$ by applying EMA (Eq. (11)).
            end for
            Calculate the prompted feature by $g_\psi(x_i; p_i^f)$.
            Obtain the target model's logits $f_{\phi_{t_i}^f}(g_\psi(\cdot))$
            Calculate the Cross-Entropy Loss $\mathcal{L}_{CE}$.
            Calculate the Input Reconstruction Loss $\mathcal{L}_{IR}$ (Eq. (9)).
            Calculate the Prompt Reconstruction Loss $\mathcal{L}_{PR}$ (Eq. (6)).
            Calculate the Total Loss $\mathcal{L}$ (Eq. (10)).
            Update $\phi_{t_i}^f$ by minimizing the Total Loss $\mathcal{L}$
        end for
    end for
end for

the complexity of foundation model for our computational resources. Nevertheless, this issue should not affect the rigor of our finding because our algorithm isn't sensitive against variations of random seeds, i.e., standard deviation is small. In addition, the macro F1 score is reported rather than the accuracy because it is more accurate than accuracy in the case of class imbalance. The hyper-parameters of consolidated algorithms are selected as per their official setting. Nonetheless, the grid search is applied when their performances are surprisingly poor.

6.3. Numerical Results

Table 2 reports the numerical results of all consolidated algorithms in the HAR dataset. CPFM (MS) denotes the average numerical results of CPFM across three random seeds while CPFM only shows the CPFM numerical results in the first seed, i.e., other algorithms are only executed under the first seed. It is clearly seen that CPFM beats other algorithm with notable margins, i.e., 11% gap to DINE in the second place as shown in Table 2. Other algorithms perform poorly confirming the challenge of time-series domain adaptation possessing unique spatio-temporal characteristics. Numerical results in the SSC dataset are tabulated in Table 3. It is seen that our algorithm, CPFM, outperforms other algorithms with significant margins, i.e., over 6% margin to BETA in the second place. Note that all consolidated algorithms are structured under the foundation model to ensure fair comparisons. As with other two cases, CPFM is also superior to its competitors with at least 2% margin to DINE in the second place in the MFD dataset as shown in the Table 4. Unfortunately, other algorithms don't perform well with over 10% gap to CPFM and DINE. This finding confirms the advantage of CPFM over the prior arts in the black-box time-series domain adaptation. In addition, direct applications of black box domain adaptation algorithms designed for vision applications to the time-series cases are not successful. That is, the time-series domain call for special treatments to cope with their spatio-temporal natures. On the other hand, the black box domain adaptation is challenging because it relies only on the API of the source model for domain adaptation. i.e, no source data nor pretrained weights are offered for domain adaptations. The underlying rationale behind higher F1 score of CPFM than other consolidated algorithms lies in the prompt tuning strategy under the dual branch network structure coupled with the dual reconstruction learning phase in both prompt and input levels assuring distinct prompts to be learned while implicitly adapting to the structure of the target domain, i.e., the input reconstruction phase learns the underlying patterns of the target domain with the absence of any labeled samples.

6.4. Multi-Source Domain Cases

We discuss the advantage of our algorithm, CPFM, under the multi-source problems. CPFM is configured under three and five source domains respectively and tested with the HAR and SSC datasets. Numerical results of the HAR dataset are reported in Table 5 while numerical results of the SSC dataset are tabulated in Table 6. It is observed that the performance of CPFM improves steadily when using 1 source domain to 5 source domains. CPFM attains around 10% improvements from 1 source to 3 sources and 11% improvements from 1 source to 5 sources in realm of the HAR dataset. The same finding takes place for the SSC dataset, i.e., 3% improvements from 1 source to 3 sources and 3.1% improvements from 1 source to 5 sources. This result confirms the efficacy of the multi-source domain strategy of CPFM using the normalized Shannon entropy and the momentum update rule. That is, such strategy enables complementary information to be mined while mitigating detrimental impact of unrelated source domains.

6.5. Ablation Study

We discuss the ablation study to verify the advantage of each learning module of CPFM where our numerical results in the HAR dataset are displayed in Table 7. We start from the efficacy of the prompt tuning strategy. That is, CPFM discards the prompt and only adjusts the classification head for domain adaptation. The absence of the prompts for domain adaptations deteriorates the performance of CPFM by about 9%. This finding confirms the advantage of our prompt tuning strategy for domain adaptations where it guides the representations of the foundation model to adapt to a downstream task. The advantage of the input reconstruction loss (9) is tested. It is perceived that CPFM loses around 2% in the MF1-score without the input reconstruction mechanism. This module plays a vital role in CPFM where it models the structure of the target domain without any labels. That is, it performs an implicit domain adaptation step. We also study the effect of the prompt reconstruction mechanism (6) to CPFM. The absence of prompt reconstruction strategy brings down the performance of CPFM by over 2%. The prompt reconstruction strategy is crucial because it underpins the creation of distinct prompts under the dual branch network structure. Last but not least, we also investigate the performance of CPFM with the naive averaging strategy in the multi-source domain adaptation phase in which Table 8 exhibits our numerical results in the HAR dataset. That is, the transferability weight is set uniformly for every source domain. This modification results in significant performance drops of CPFM for both 3 and 5 source domains confirming the advantage of our normalized entropy and momentum update strategy in the multi-source domain adaptation problems.

6.6. TSNE Analysis

Fig. 3 and 4 visualize the TSNE plots of the network branch 1 before and after the training process respectively for the SSC dataset while Fig. 5 and Fig. 6 depict the TSNE plots of the network branch 2 before and after the training process respectively for the SSC dataset. As CPFM is built upon the dual-branch network structure, the TSNE plots encompass the embedding of each network branch. It is seen that even before the training process begins, the foundation model enjoys generalizable features. These features turn to be discriminative after the training process because the

Table 2
Five HAR cross-domain scenarios results in terms of MF1 score. MS means Multi-Seed

Method	2 → 11	6 → 23	7 → 13	9 → 18	12 → 16	AVG
BETA	17.8	17.93	18.26	16.61	20.75	18.27
CoMDA	4.8	5.26	5.37	5.56	4	4.99
DINE	36.92	33.15	47.47	23.87	21.26	32.53
RFC	18.53	15.33	9.09	16.18	12.58	14.34
SEAL	7.44	9.43	17.21	10.29	5.81	10.04
CPFM	**56.08**	**37.94**	**59.81**	**34.27**	42.66	**46.15**
CPFM (MS)	48.60±10.89	36.89±1.37	56.77±11.27	32.22±2.00	**44.91±2.31**	44.08

Table 3
Five SSC cross-domain scenarios results in terms of MF1 score. MS means Multi-Seed

Method	0 → 11	12 → 5	7 → 18	16 → 1	9 → 14	AVG
BETA	23.88	22.68	53.97	49.81	37.67	37.6
CoMDA	13.67	20.42	20.97	17.45	13.36	17.17
DINE	24.85	24.7	40.96	37.38	14.88	28.55
RFC	18.39	21.16	27.55	9.38	27.05	20.71
SEAL	12.87	19.91	14.11	26.06	24.71	19.53
CPFM	**32.17**	**35.41**	55.47	**58.04**	37.98	**43.81**
CPFM (MS)	31.87±0.93	31.71±5.84	**55.57±1.20**	56.86±3.56	**42.97±7.06**	43.80

Table 4
Five MFD cross-domain scenarios results in terms of MF1 score. MS means Multi-Seed

Method	0 → 1	1 → 2	3 → 1	1 → 0	2 → 3	AVG
BETA	20.83	20.83	20.83	5.58	20.83	17.78
CoMDA	5.56	5.56	5.56	5.56	5.56	5.56
DINE	20.81	32.79	**75.17**	**66.65**	32.37	45.56
RFC	**34.67**	35.55	20.83	20.83	52.39	32.85
SEAL	20.83	20.83	20.83	5.58	20.83	17.78
CPFM	20.81	**52.86**	55.31	51.06	**55.57**	**47.12**
CPFM (MS)	20.81±0	52.62±0.22	55.38±0.12	47.32±5.74	51.65±4.38	45.56

Table 5
Five HAR multi-source cross-domain scenarios results in terms of MF1 score. 1S for single-source, 3S and 5S for three and five sources respectively

Method	→ 11	→ 23	→ 13	→ 18	→ 16	AVG
CPFM (1S)	56.08	37.94	59.81	34.27	42.66	46.15
CPFM (3S)	37.91	65.25	54.4	46.39	72.45	55.28
CPFM (5S)	35.78	72.51	67.39	51.77	52.42	55.97

Table 6
Five SSC multi-source cross-domain scenarios results in terms of MF1 score. 1S for single-source, 3S and 5S for three and five sources respectively

Method	→ 11	→ 5	→ 18	→ 1	→ 14	AVG
CPFM (1S)	32.17	35.41	55.47	58.04	37.98	43.81
CPFM (3S)	35.41	41.38	55.33	49.43	48.54	46.02
CPFM (5S)	28.62	44.16	58.29	54.9	44.77	46.15

Table 7
Ablation study in HAR dataset.

Method	2 → 11	6 → 23	7 → 13	9 → 18	12 → 16	AVG
CPFM	56.08	37.94	59.81	34.27	42.66	46.15
CPFM (w/o Prompt)	47.45	29.58	37.28	32.83	41.13	37.65
CPFM (w/o Input Recons)	42.85	38.34	68.72	32.62	41.18	44.74
CPFM (w/o Prompt Recons)	51.82	39.18	61.3	33.41	35.23	44.19

Table 8
Comparisson of multi-source domain cases in HAR dataset. 3S and 5S for three and five sources respectively

Method	→ 11	→ 23	→ 13	→ 18	→ 16	AVG
CPFM (3S)	37.91	72.44	65.25	46.39	54.4	55.28
CPFM (5S)	35.78	72.51	67.39	51.77	52.42	55.97
CPFM NaiveAvg (3S)	32.59	58.25	73.96	42.99	54.93	52.54
CPFM NaiveAvg (5S)	39.86	57.18	62.32	52.79	52.74	52.98

prompts are learned and the backbone networks are frozen. That is, the use of adjustable prompts enhances the embedding qualities. Both network branch 1 and 2 induce the same embeddings before the training process because of the same prompts. The advantage of the prompt reconstruction strategy can be also viewed here where it generate distinct prompts inducing complementary information, i.e., the embeddings of the two branches are distinguishable after the training process, thus implying distinct prompts because the backbone network is unchanged during the training process. It is also perceived that the source and target samples are mapped closely after the training process.

7. Conclusion

This paper studies the problem of black box time-series domain adaptation (BBTSDA) and proposes a novel algorithm, termed cross-prompt foundation model (CPFM) for solving the BBTSDA problem. CPFM is built upon a time-series foundation model coupled with the prompt tuning strategy. We put forward the notion of the dual-branch network structure where a unique prompt is attached to each network branch to generate complementary information. This strategy is supported with the prompt reconstruction strategy to produce distinct prompts while the input reconstruction strategy functions as the implicit domain adaptation step by modeling the target domain directly without any labels. CPFM is also adept at the multi-source domain adaptation case using the idea of normalized entropy and momentum update technique. Our numerical results confirm the advantage of CPFM over prior arts with noticeable margins in three datasets of different application domains. The performance of CPFM steadily increases with the number of source domains while the ablation study bears out the positive impact of each learning module. Our study still assumes the closed-set scenario where the source and target domains share identical label space. Our future study will be devoted to explore the category shift problem in the time-series domain adaptation.

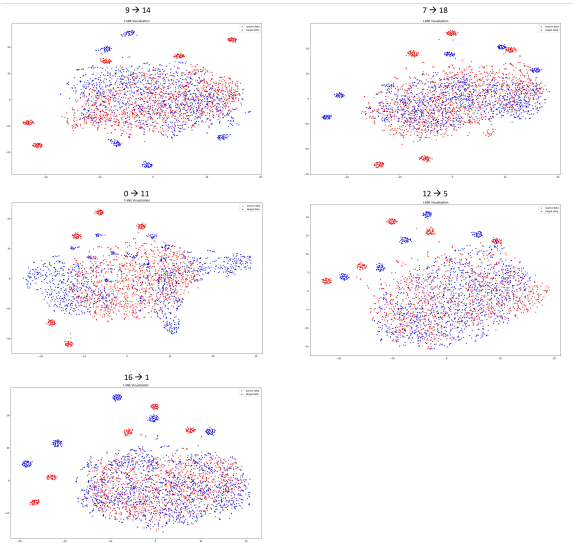

Figure 3: T-SNE of SSC Dataset by network branch 1 before training

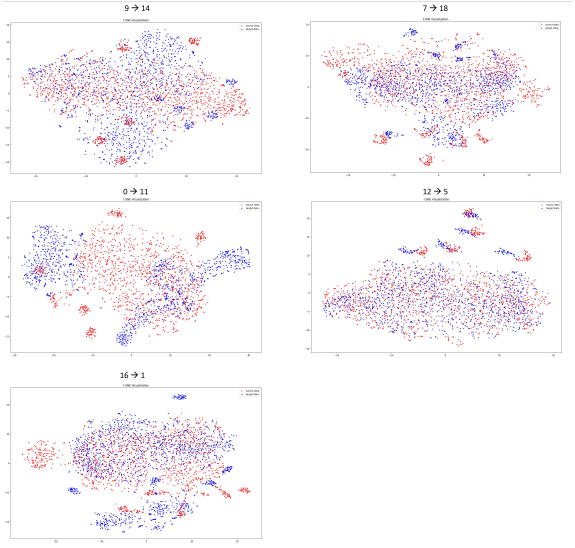

Figure 4: T-SNE of SSC Dataset by network branch 1 after training

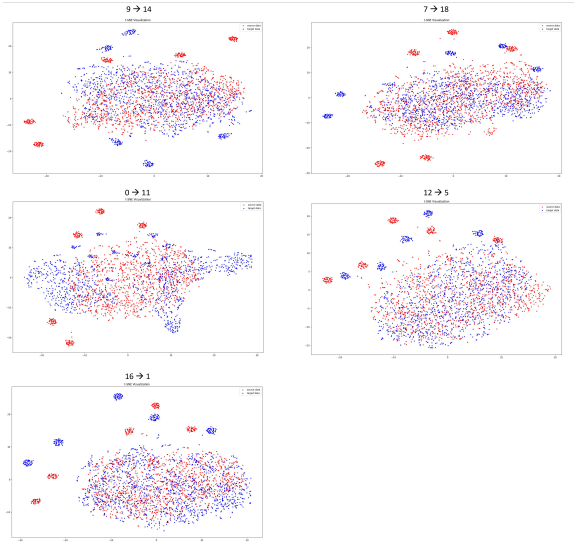

Figure 5: T-SNE of SSC Dataset by network branch 2 before training

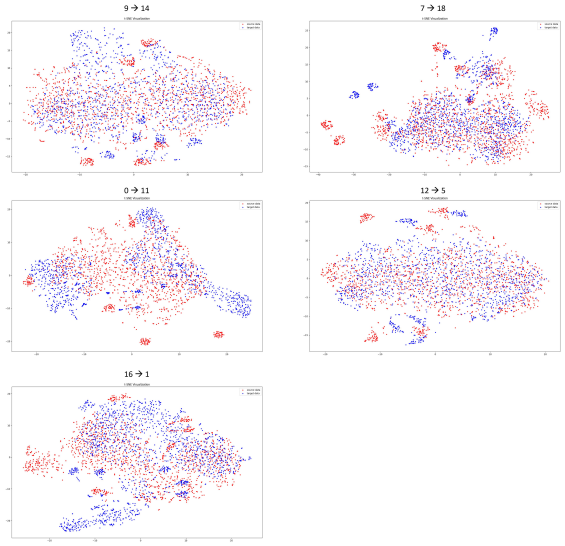

Figure 6: T-SNE of SSC Dataset by network branch 2 after training

References

Cai, R., Chen, J., Li, Z., Chen, W., Zhang, K., Ye, J., Li, Z., Yang, X., Zhang, Z., 2020. Time series domain adaptation via sparse associative structure alignment, in: AAAI Conference on Artificial Intelligence. URL: https://api.semanticscholar.org/CorpusID:229349290.

Chen, D., Wang, D., Darrell, T., Ebrahimi, S., 2022a. Contrastive test-time adaptation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) , 295–305URL: https://api.semanticscholar.org/CorpusID:248366600.

Chen, H., Wu, Z., Jiang, Y.G., 2022b. Multi-prompt alignment for multi-source unsupervised domain adaptation. ArXiv abs/2209.15210. URL: https://api.semanticscholar.org/CorpusID:252668638.

Furqon, M., Pratama, M., Liu, L., Habibullah, H., Doğançay, K., 2024a. Mixup domain adaptations for dynamic remaining useful life predictions. Knowledge-Based Systems URL: https://api.semanticscholar.org/CorpusID:269005807.

Furqon, M.T., Pratama, M., Shiddiqi, A.M., Liu, L., Habibullah, H., Doğançay, K., 2024b. Time and frequency synergy for source-free time-series domain adaptations. ArXiv abs/2410.17511. URL: https://api.semanticscholar.org/CorpusID:273532619.

Ganin, Y., Lempitsky, V.S., 2014. Unsupervised domain adaptation by backpropagation, in: International Conference on Machine Learning. URL: https://api.semanticscholar.org/CorpusID:6755881.

Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., Dubrawski, A., 2024. Moment: A family of open time-series foundation models. ArXiv abs/2402.03885. URL: https://api.semanticscholar.org/CorpusID:267500205.

Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I.W.H., Sugiyama, M., 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels, in: Neural Information Processing Systems. URL: https://api.semanticscholar.org/CorpusID:52065462.

Jahan, C.S., Savakis, A.E., 2023. Curriculum guided domain adaptation in the dark. IEEE Transactions on Artificial Intelligence 5, 2604–2614. URL: https://api.semanticscholar.org/CorpusID:260378920.

Jiao, J., Zhang, T., Li, H., Liu, H., Lin, J., 2025. Source-free black-box adaptation for machine fault diagnosis. IEEE Transactions on Industrial Informatics URL: https://api.semanticscholar.org/CorpusID:275874232.

Jin, X., Park, Y., Maddix, D.C., Wang, B., Yan, X., 2021. Domain adaptation for time series forecasting via attention sharing, in: International Conference on Machine Learning. URL: https://api.semanticscholar.org/CorpusID:235421595.

Kang, G., Jiang, L., Yang, Y., Hauptmann, A., 2019. Contrastive adaptation network for unsupervised domain adaptation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) , 4888–4897URL: https://api.semanticscholar.org/CorpusID:57572938.

Karim, N., Mithun, N.C., Rajvanshi, A., Chiu, H.P., Samarasekera, S., Rahnavard, N., 2023. C-sfda: A curriculum learning aided self-training framework for efficient source free domain adaptation. ArXiv abs/2303.17132. URL: https://api.semanticscholar.org/CorpusID:257833516.

Li, R., Jiao, Q., Cao, W., Wong, H.S., Wu, S., 2020. Model adaptation: Unsupervised domain adaptation without source data. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) , 9638–9647URL: https://api.semanticscholar.org/CorpusID:219979590.

Liang, J., Hu, D., Feng, J., 2020. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation, in: International Conference on Machine Learning. URL: https://api.semanticscholar.org/CorpusID:211205159.

Liang, J., Hu, D., Feng, J., He, R., 2021. Dine: Domain adaptation from single and multiple black-box predictors. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) , 7993–8003URL: https://api.semanticscholar.org/CorpusID:244800743.

Litrico, M., Bue, A.D., Morerio, P., 2023. Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation. ArXiv abs/2303.03770. URL: https://api.semanticscholar.org/CorpusID:257378054.

Liu, C., Zhou, L., Ye, M., Li, X., 2022. Self-alignment for black-box domain adaptation of image classification. IEEE Signal Processing Letters 29, 1709–1713. URL: https://api.semanticscholar.org/CorpusID:251473244.

Liu, X., Xi, W., Li, W., Xu, D., Bai, G., Zhao, J., 2023. Co-mda: Federated multisource domain adaptation on black-box models. IEEE Transactions on Circuits and Systems for Video Technology 33, 7658–7670. URL: https://api.semanticscholar.org/CorpusID:258776491.

Lü, S., Kang, M., Li, X., 2024. Alleviating imbalanced pseudo-label distribution: Self-supervised multi-source domain adaptation with label-specific confidence. Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence URL: https://api.semanticscholar.org/CorpusID:271508637.

Raffel, C., Shazeer, N.M., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J., 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. 21, 140:1–140:67. URL: https://api.semanticscholar.org/CorpusID:204838007.

Ragab, M., Eldele, E., Chen, Z., Wu, M., Kwoh, C., Li, X., 2021. Self-supervised autoregressive domain adaptation for time series data. IEEE transactions on neural networks and learning systems PP. URL: https://api.semanticscholar.org/CorpusID:244729392.

Ragab, M., Eldele, E., Wu, M., Foo, C.S., Li, X., Chen, Z., 2023. Source-free domain adaptation with temporal imputation for time series data. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining URL: https://api.semanticscholar.org/CorpusID:259937854.

Ren, L., Cheng, X., 2023. Single/multi-source black-box domain adaption for sensor time series data. IEEE transactions on cybernetics PP. URL: https://api.semanticscholar.org/CorpusID:261073969.

Wang, L., Yang, X., Tan, H., Bai, X., Zhou, F., 2023. Few-shot class-incremental sar target recognition based on hierarchical embedding and incremental evolutionary network. IEEE Transactions on Geoscience and Remote Sensing 61, 1–11. URL: https://api.semanticscholar.org/CorpusID:257147398.

Wang, Z., Zhang, Z., Ebrahimi, S., Sun, R., Zhang, H., Lee, C.Y., Ren, X., Su, G., Perot, V., Dy, J.G., Pfister, T., 2022. Dualprompt: Complementary prompting for rehearsal-free continual learning. ArXiv abs/2204.04799. URL: https://api.semanticscholar.org/CorpusID:248085201.

Wilson, G., Doppa, J.R., Cook, D.J., 2020. Multi-source deep domain adaptation with weak supervision for time-series sensor data. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining URL: https://api.semanticscholar.org/CorpusID:218863107.

Xia, M., Zhao, J., Lyu, G., Huang, Z., Hu, T., Chen, G., Wang, H., 2024. A separation and alignment framework for black-box domain adaptation, in: AAAI Conference on Artificial Intelligence. URL: https://api.semanticscholar.org/CorpusID:268708817.

Xiao, S., Ye, M., He, Q., Li, S., Tang, S., Zhu, X., 2024. Adversarial experts model for black-box domain adaptation, in: ACM Multimedia. URL: https://api.semanticscholar.org/CorpusID:273645973.

Yang, J., Peng, X., Wang, K., Zhu, Z.H., Feng, J., Xie, L., You, Y., 2022. Divide to adapt: Mitigating confirmation bias for domain adaptation of black-box predictors. ArXiv abs/2205.14467. URL: https://api.semanticscholar.org/CorpusID:249191952.

Zhang, J., Huang, J., Jiang, X.Q., Lu, S., 2023. Black-box unsupervised domain adaptation with bi-directional atkinson-shiffrin memory. 2023 IEEE/CVF International Conference on Computer Vision (ICCV) , 11737–11748URL: https://api.semanticscholar.org/CorpusID:261214533.

Zhang, S., Shen, C., Lü, S., Zhang, Z., 2024. Reviewing the forgotten classes for domain adaptation of black-box predictors, in: AAAI Conference on Artificial Intelligence. URL: https://api.semanticscholar.org/CorpusID:268696714.

Zhao, Y., Feng, S., Li, C., Song, R., Liang, D., Chen, X., 2023. Source-free domain adaptation for privacy-preserving seizure prediction. IEEE Transactions on Industrial Informatics URL: https://api.semanticscholar.org/CorpusID:260412519.